THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 58, PART 6 1 November 1946 No. 330

RUTHERFORD: LIFE AND WORK AFTER THE YEAR 1919, WITH PERSONAL REMINISCENCES OF THE CAMBRIDGE PERIOD

BY J. D. COCKCROFT, F.R.S.

Second Rutherford Memorial Lecture, delivered 8 *February* 1946

§ 1. INTRODUCTION

THE Council of the Physical Society in establishing the Rutherford Memorial lecture directed that the first two of these should be reminiscent and personal. This would, I feel, have met with approval from Rutherford, whose love of personal reminiscences, and particularly of indiscretions, was a well known characteristic; so I shall begin by saying what qualifications I have for speaking of Rutherford and his laboratory in this personal way.

I first met Rutherford in 1914 in Manchester, where I had gone to sit at the feet of that most charming of mathematicians, Horace Lamb. As some light relief I attended the first-year lectures in physics. These lectures increased steadily in noisiness until one day the storm broke, and Rutherford was brought in to restore order. I still remember the immediate impression that here was a great man who was not going to stand any nonsense; thereafter the lectures were delivered by Rutherford in perfect quiet except for the applause which greeted the beautiful demonstrations of Kay, the laboratory steward. This was the first year of war, and my contact with Rutherford was broken for about seven years. I was then persuaded by Miles Walker in 1922 to take up again the study of mathematics in Cambridge, and was given the best of introductions to Cambridge, a letter to Rutherford. I remember going to see him in the old Maxwell Wing of the laboratory and finding him sitting, as he so often did, on a stool. He received me very kindly, and gave me authority to devote such time as I could spare from mathematics to work in the advanced practical class where, under Thirkill and Appleton, I learnt the elements of laboratory arts. After two years of the Tripos I was accepted as a member of the laboratory, and set to work in Chadwick's introductory training course. I worked first with Joseph Boyce on building and using the favourite tool of the Cavendish, the gold-leaf electroscope. After that we blew and built with our own hands a fine Macleod gauge—and were horrified by destroying it in the classical manner just as Rutherford walked in at the door. He was sympathetic, but explained that he did not keep students who had too many accidents. This dislike of sins against apparatus was one of Rutherford's well marked characteristics.

By this time, in 1924, Rutherford's Cambridge Research School was well established. The story of his second coming to Cambridge has been told in Eve's Biography: after the first approach from the University there was a typical Rutherford letter to J. J. Thomson asking for a clear mutual understanding with regard to the laboratory and research students. These negotiations went on for some time, helped by Sir Joseph Larmor acting as intermediary, and finally the conclusions were embodied in a document which formed the working agreement between Thomson and Rutherford for their respective spheres of influence in the Cavendish Laboratory. This agreement is reproduced in figure 1, whilst figure 2 shows the two signatories in later years. As a result of this agreement "J. J." retained the ground floor of the wing of the Cavendish built by Lord Rayleigh, a space always referred to as "The Garage". There he worked with his personal assistant Everett, who taught us glass blowing and sold to us boxes of "Everett wax". Every day at 1 o'clock or at the end of the day we would see J. J. standing at his little desk looking over the results of Everett's experimenting, or peering through his spectacles at the maze of glass tubing, which formed his experiment. He was a much-loved and venerable figure, always giving us a cheerful grin on passing, and always full of talk on topics other than physics.

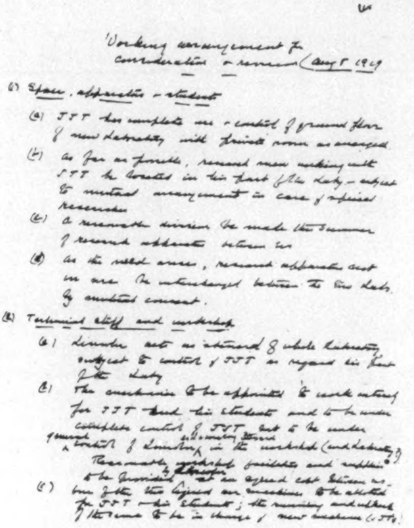
Within the J. J. domain worked Aston, his research on isotopes then coming into full flower. During the post-war years he was able to separate the isotopes of neon, chlorine, mercury and lithium and to build thereafter the first of the precision spectrographs, shown in figure 3.

§2. BEFORE 1930

Rutherford's interests during the first decade of his work at Cambridge followed three main lines. First came his work on the transmutation of the lighter elements. This work was a continuation of that started in Manchester during the war years. An observation of Marsden in 1914 had shown that long-range particles were produced when alpha particles were fired into hydrogen gas. Working with Kay, the Manchester Steward, Rutherford showed that these particles were hydrogen nuclei. He went further and showed that similar particles were produced when alpha particles were fired into nitrogen. He concluded that a transmutation of nitrogen had been achieved. These experiments, published in the *Philosophical Magazine* in 1919, were a fitting climax to Rutherford's work in Manchester.

The work was taken up at once at Cambridge in Maxwell's old research room. Figure 4 shows the apparatus which he used with Chadwick. The alpha particles from the radioactive source passed through a brass chamber filled with nitrogen or other gases. The protons ejected were detected by the simplest of devices, a zinc sulphide screen. Between the screen and the chamber, absorbing screens of different thicknesses could be inserted.

Figure 5 shows Rutherford with his apparatus. Every day round about 4 o'clock periods for counting were arranged and two of the research students with good eyes were brought in to help. They sat in the darkened room for half an hour, and over cups of tea listened to Rutherford talking about all things



Working arrangement for consideration and revision (Aug 8 1919

(1) Space, apparatus and students

(a) J.J.T. has complete use and control of ground floor of new laboratory with private room as arranged.

(b) As far as possible, research men working with J.J.T. be located in this part of the Laby — subject to mutual arrangement in case of special researches

(c) A reasonable division be made this summer of research apparatus between us

(d) As the need arises, research apparatus not in use be interchanged between the two Labs. by mutual consent.

(2) Technical staff and workshop

(a) Lincoln acts as steward of whole laboratory subject to control of J.J.T. as regards his part of the Laby

(b) One mechanic to be appointed to work entirely for J.J.T. and his students and to be under complete control of J.J.T., but to be under general control of Lincoln [illegible] in the workshop (and Laboratory?)

Reasonable workshop facilities and supplies to be provided by [illegible] at an agreed cost between us.

(c) One of the three lathes and machines to be allotted for J.J.T. and his students; the running and upkeep of the same to be in charge of new mechanic (i.e. J.J.T.'s).

Figure 1.

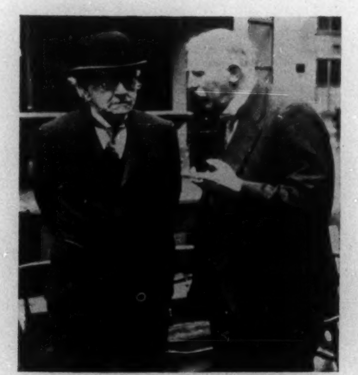


Figure 2.

Figure 3.

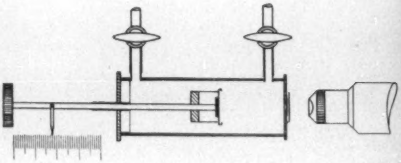


Figure 4.

Figure 5.

under the sun. Thereafter when Crowe had produced the source they took turns of a minute each to look through the microscope and to count the faint proton scintillations.

Mr. G. R. Crowe, who was Rutherford's assistant, gave to Eve a lively description of the work done at that time. He said, "He was shooting protons out of light atoms by means of alpha particles; carbon would not play". The next day was the turn of aluminium, and on some theoretical ground Rutherford thought the protons from aluminium would have a high velocity and a long range. "Now, Crowe, have some mica absorbers ready tomorrow with a stopping power of 50 cm."——"Yes, sir". "Now, Crowe, put in a 50 cm. screen"——"Yes, sir". "Why don't you do what I tell you; put in a 50 cm. screen"——"I have, sir". "Put in 20 more"——"Yes, sir". "Why the devil don't you do what I tell you; I said 20 more"——"I did, sir". "There's some damned contamination; put in two 50's"——"Yes, sir". "Ah, it's all right; that's stopped 'em. Crowe, my boy, you're always wrong until I've proved you right. Now we'll find their exact range".

Ellis, in his Obituary of Rutherford,* has described how firmly Rutherford kept control of the experiments in the darkened room. As the scintillations were recorded, Rutherford would compare them with what he expected and translate them at once into the final deductions. During the years up to 1925, transmutation of the light elements up to potassium was studied, and the ranges of the disintegration protons determined; the identity of the disintegration particles was proved with characteristic thoroughness by measuring their deflection in a magnetic field.

During this period Blackett embarked on his courageous attempt to get a Wilson chamber photograph of the nitrogen disintegration. He took photographs of hundreds of thousands of tracks and was eventually rewarded (figure 6) by seeing the tracks of disintegration protons. The photograph shows a long proton track shooting downwards. An alpha particle has entered a nitrogen nucleus, ejecting a proton and leaving oxygen behind, producing the transmutations ${}^{14}_{7}\mathrm{N} + {}^{4}_{2}\mathrm{He} \rightarrow {}^{17}_{8}\mathrm{O} + {}^{1}_{1}\mathrm{H}$.

This was the first direct proof that the alpha particle is captured in the transmutation.

The second main line of experimenting in the laboratory in this decade was a study of the fields of force round atomic nuclei. The first simple scattering experiments of Rutherford and Marsden in 1907 had led to Rutherford's fundamental concept of the nuclear atom. These experiments were continued in the post-war years by the more refined experiments of Chadwick and Bieler. They shot alpha particles of varying velocity through gases and foils, measured the angular distribution and compared it with that to be expected if the inverse square law held. They were able to show that near heavy nuclei, such as gold, the inverse square law held up to the limits of penetration of the fastest alpha particles available. But with the lighter elements they found very marked deviations. These experiments presented Rutherford with the paradox that the potential barrier round the heavy nuclei must rise well beyond 10 million volts, whereas alpha particles come out of these nuclei at energies between

---

* *Proc. Phys. Soc.*, **50,** 463 (1938).

4 and 8 million volts. The alpha particles do not therefore escape with the full energy of the potential field.

The solution to this problem was provided by a visitor to the laboratory; a young and imaginative theoretical physicist from Leningrad, Dr. Gamow. Gamow saw that the solution was provided by the new wave mechanics which was then being developed by Schrödinger and others. On the new ideas it was evident that the alpha particles could leak through the potential barriers of the nuclei, and so need not have the full energy of the field. In this way, nuclear physicists were given a new and most fruitful model to guide their researches.

The third main line of the laboratory during this period was the continuation of Rutherford's Manchester and McGill interests in the radiations from radioactive substances. Ellis applied the magnetic focusing method to the study of the gamma rays from the radioactive substances. He was able to show that the gamma rays resulted from transitions between nuclear energy levels.

These gamma-ray observations were linked in the laboratory with a study of the alpha-particle spectra. We built for Rutherford a special annular magnet whereby he could study the fine structure of these lines. The nuclear energy levels so determined agreed in general with those found by Ellis's gamma-ray work.

(At this point, the lecturer had records played in which Rutherford himself was heard speaking of this work.)

§ 3. AFTER 1930

The second decade of Rutherford's work in Cambridge opened with an intensive development of new laboratory techniques. In various remote corners and cellars, Wynn-Williams and his co-workers had been developing methods for the electrical recording of particles. In other parts of the laboratory, Feather and Nimmo, E. J. Williams, Terroux and others were building cloud chambers, spreading the influence of C. T. R. Wilson throughout the laboratory. Allibone came in from Metropolitan-Vickers, bringing with him a fierce Tesla coil which produced 500,000-volt sparks, to the annoyance of the Corpus dons across the way, and proceeded to build in the laboratory a high-voltage tube for the acceleration of electrons.

Rutherford's Royal Society Presidential address of 1927 had urged the development of sources of atoms and electrons with an energy far transcending that of the alpha and beta particles from radioactive matter. Allibone's 300-kilovolt electrons were the first step on the way.

Walton came to us from Trinity College, Dublin, and at Rutherford's suggestion started to work on a new method of accelerating electrons by spinning them round in the electric field produced by a changing magnetic field. This was the principle of the betatron. Walton was able to develop the theory of the stability of the orbits which has since been used as the basis of design of successful betatrons. The experimental problem was attacked with very limited resources and Walton was not able to push the work through to a successful conclusion.

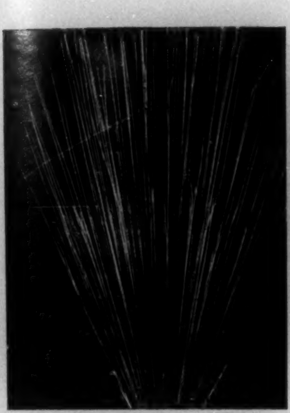


Figure 6.

Figure 8.

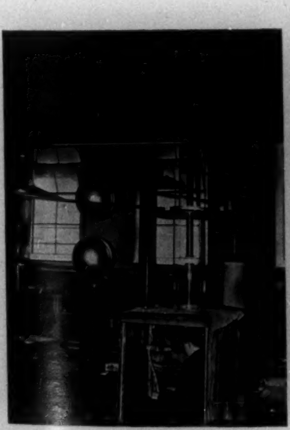


Figure 7.

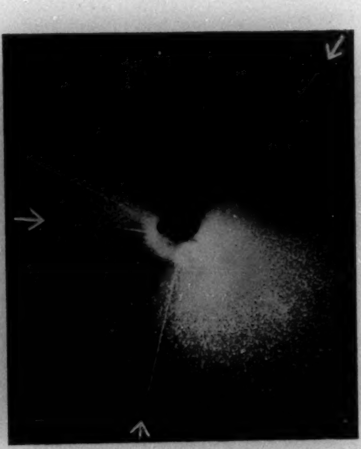


Figure 9.

Figure 10 (*a*).

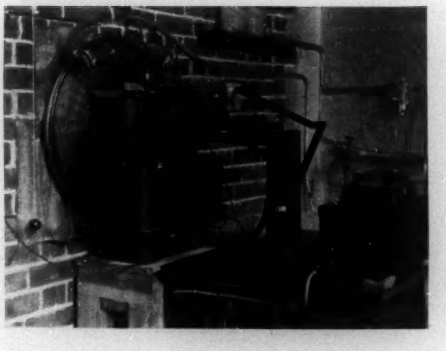


Figure 10 (*b*).

About the same time I started to build the first high-voltage apparatus for the acceleration of protons. I was led to do this by the predictions of Gamow's theory that protons of a few hundred kilovolts should penetrate the barriers of light nuclei. These predictions were submitted to Rutherford and he encouraged me to go ahead.

The fruit of all this development of technique came in that very wonderful year of 1932. First of all came the dramatic discovery of the neutron by Chadwick. For some time the laboratory had been interested in the radiations emitted when various light elements such as beryllium were bombarded with alpha particles. The radiations from beryllium were at first thought to be gamma rays, but did not behave in the expected manner, and for several months experiments were going on to try and find out what the radiation was. The announcement by Curie and Joliot of the projection of hydrogen nuclei by this radiation gave the clue to Chadwick. By a series of very rapid experiments in which the projection of different nuclei was studied, he was able to establish without doubt that the recoil protons were not due to a Compton effect, but resulted from the impact of material particles of unit mass and zero charge.

Very swiftly, Dee and Feather were set to work with their expansion chambers to look for the tracks of the projected particles. Dee found the tracks of proton and other recoils and Feather discovered the transmutation of nitrogen by neutrons, the first of the transmutations by this new particle.

Whilst this was going on, Walton and I had been completing our apparatus for the acceleration of protons. We had everything stuck together with plasticene by January 1932, and started to experiment with our proton beam, which passed out from the tube and travelled over 1 cm. in air. Figure 7 shows the apparatus. Rutherford was a little impatient at this stage, being very anxious to see the results. We put in a lithium target and a scintillation screen and were rewarded at once by seeing the bright scintillations characteristic of alpha particles. (In figure 7 Walton is seen sitting in the box with the recording apparatus used for looking for particles from disintegration.) We called in Rutherford at once and he was very enthusiastic. A few days later he announced the results with great gusto at a discussion of the Royal Society.

Rutherford's sound strategical sense was typified by his immediate diversion of P. I. Dee from his work with C. T. R. Wilson to join with Walton and myself in the study of the disintegrations by the cloud-chamber method. In figure 8 is a very beautiful picture taken by Dee in which the two alpha particles from one of these transmutations are seen travelling in opposite directions, whilst figure 9 is another Wilson chamber photograph showing the three alpha particles (joined by arrows) coming out of the transmutation of boron. The boron unites with a proton and splits up into three alpha particles.

Following on this, Oliphant was diverted from his work on positive ions to apply his technique to the production of strong proton beams for these experiments. Rutherford had to have an apparatus of his own to work with. Figure 10 shows the low-voltage apparatus built by Oliphant and Kempton in which 100 microamperes of protons were accelerated by about 200,000 volts.

About the same time in 1933 Rutherford received from G. N. Lewis the first sample of heavy hydrogen to reach Europe. Within a few days the sample was converted into deuterium gas, which Rutherford guarded with the most jealous care. He turned at once with Oliphant to do experiments with deuterons and soon discovered the transmutation of deuterium by deuterons. They found that two reactions occurred, $D + D = {}^{1}H + {}^{3}H$ or ${}^{3}He + {}^{1}_{0}n$, leading to the new elements ${}^{3}H$ and ${}^{3}He$. Two proton groups were discovered in the first reaction, and a short-range alpha-particle group in the second.

Rutherford and Oliphant's speculative interpretation of the reactions was confirmed by Dee's Wilson-chamber photography, Dee being lent for a few days a few cc.s of the precious gas. Figure 11 shows the tracks from the reaction ${}^{2}_{1}D + {}^{2}_{1}D = {}^{1}_{1}H + {}^{3}_{1}H$. The track on the left is a proton; the track on the right is the new kind of hydrogen of mass 3, a hydrogen which we now know to be radioactive, with a very long life.

One other discovery of major importance was made during this same period. During 1932 we had a visit from Millikan, who brought with him some very remarkable cosmic-ray photographs taken by Anderson which gave the first indications of positive electrons. Immediately after that, Blackett and Occhialini turned on their Wilson chamber to the search. Introducing the principle of counter control, they very soon obtained some remarkable photographs, showing pairs of positive and negative electrons, and in some cases showers of particles. Figure 12 shows one of these photographs.

This was followed up by Chadwick, Blackett and Occhialini's discovery of the production of electron pairs by high-energy gamma radiations, and finally, to close this sequence of exciting discoveries, there came Chadwick and Goldhaber's discovery of the photoelectric disintegration of deuterium by gamma rays.

The discovery of artificial radioactivity was missed in the laboratory, largely because we did not in general work with Geiger counters, and were looking for particles either with scintillation screens or with counters which would not respond to beta rays. When Curie-Joliot announced the discovery of the production of artificial radioactivity of alpha particle bombardment, Walton and I were able to borrow a Geiger counter equipment from Bainbridge and to show that protons could produce artificial radioactivity in carbon.

§4. THE DEVELOPMENT OF LOW-TEMPERATURE WORK IN CAMBRIDGE

The development of low-temperature work in the Cavendish Laboratory by Kapitza is one case where Rutherford backed strongly research work outside his own special interests.

Kapitza had come to England about 1922 on some Soviet Government mission. When I first met him in 1924 he was engaged in producing strong magnetic fields in small coils by discharging a home-made accumulator battery through the coil. The ground floor room of the block which is now devoted to Colloid Science resonated with the hammering of lead plates to straighten out defective

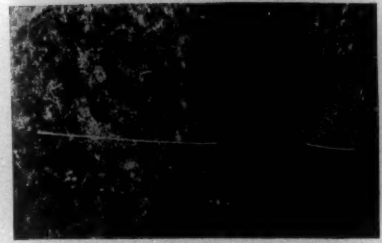


Figure 11.

Figure 13.

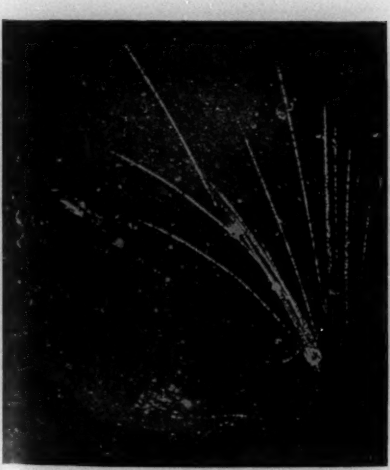


Figure 12. [With acknowledgments to Prof. P. M. S. Blackett; reproduced by courtesy of Messrs. Longmans, Green and Co., Ltd.]

Figure 14.

Figure 15.

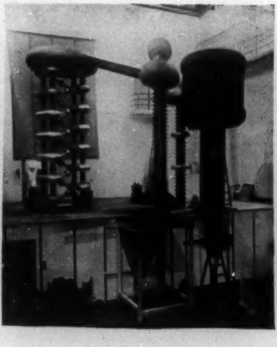


Figure 16.

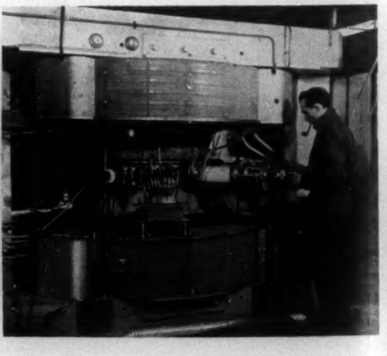


Figure 17.

members. These fields had been used to curl alpha particle tracks and to show the loss of charge at the end.

Kapitza next designed an alternator to deliver currents of 20,000 ampères on short-circuiting the machine through his coils. The D.S.I.R. provided the necessary funds, Metropolitan-Vickers designed and delivered the machine, and after many tribulations Kapitza built coils which stood up to the very high internal electromagnetic stress developed.

This led to the study of the electrical resistance of metals in strong magnetic fields. The phenomena were found to be specially interesting at low temperatures, and so a hydrogen liquefier was built and installed in buildings wrung from the reluctant chemists.

After this, Kapitza determined to liquefy helium. More space was required, and Rutherford was prevailed upon to seek the help of the Royal Society. As a result the Society provided from the Mond Fund sufficient funds to build the Royal Society Mond Laboratory. This laboratory was completed in 1933, and Kapitza built there the ingenious liquefier which for the first time eliminated the use of liquid hydrogen as an intermediary step in the liquefaction process.

Kapitza went to Russia in 1934 and did not return. After some time Rutherford entered into negotiations with the Russian Government and Kapitza's equipment was transferred to the Institute for Physical Problems in Moscow which Kapitza built. The Royal Society Mond Laboratory continued, however, to develop low-temperature work, and the discovery of new properties of liquid helium II was an exciting reward for the years spent in building up the new techniques in Cambridge.

§ 5. RUTHERFORD'S METHODS OF WORK

Finally I should say something about Rutherford's way of running the laboratory. This was very characteristic of the man. He devoted a great deal of thought to the selection of problems to be worked on by research students. Readers of Eve's *Life* will remember that in the Manchester days he used to carry in his pocket a long list of problems, and there were always plenty of spare problems.

In the later Cavendish period a list of problems was collected during the Long Vacation from the senior members of the laboratory, and Rutherford took this away with him to his country cottage and thought over it at leisure. At the beginning of the October term the problems were ready to be assigned to students after their initial training period in Chadwick's attic course.

Throughout the year, however, Rutherford was always keenly alive to new situations. I have already given you examples of the speed with which research workers were switched over to bring different techniques to bear on new problems.

With increasing duties, in London, Rutherford's tours of the laboratory become less frequent, but his visits and discussions, sitting as he used to do on a laboratory stool to go over the latest results, were always very welcome and inspiring except for those rare days when he was in a black mood. These storms of Rutherford's used to blow up very suddenly, often started by mishandling of a favourite piece of apparatus. Then he raged through the laboratory and told everybody he met the worst that could be said about their habits of work and their experiments. Some of it was not fully justified, and after the storm it was all forgotten, but the prospect of these crushing analyses did quite a bit to keep the laboratory on its toes. Rutherford told me once he had mellowed a great deal in his later years. If so, these storms in Manchester must have been really awe-inspiring spectacles !

Figures 13 and 14 respectively show Rutherford sitting on a stool talking to Kempton and Wescott and to Ratcliffe on one of his tours of the laboratory.

One of the things that Rutherford never forgave was the publication of wrong results. For these he had a very long memory. He believed in the notice written in the entrance to McGill Physics Laboratory—" Prove all things ".

In his later years Rutherford was not able to engage with his own hands in experiments, but from time to time he would try his skill. I remember Crowe telling us about an occasion when he was trying to fix a thin mica window (which is apt to be a delicate operation). His hands were shaking rather too much to make it go. He said to Crowe: " Crowe, why the devil are you shaking the table ? "——" I'm not, sir ", said Crowe. " Crowe, why the devil are you shaking the table ? "——" I'm not, sir ". " Oh, well, Crowe, you come and fix it yourself ".

Rutherford did not believe in committees for running the laboratory. He had his own system which he called " polling the jury ". On all important matters of laboratory policy one or two senior members of the staff were summoned individually to his office and the matter was very thoroughly discussed. From the advice given to him Rutherford made up his own mind, and he was very rarely wrong.

The detailed work of the laboratory was divided up amongst the seniors. Thirkill did most of the supervision of accounts and represented the laboratory on the principal University Committees. Chadwick was responsible for the detailed supervision of the research students and for the provision of research apparatus. I looked after buildings and engineering equipment; Dee and Ratcliffe were in charge of Part II teaching and were responsible for the recommendations of students to go on to research. Very great care was devoted to the selection of these students. Experimental ability was always taken into account as well as paper results.

One of Rutherford's characteristics was his willingness to let his boys go ahead and develop new techniques and to spend quite a long time doing it—one or two years—without any appreciable interference or control. He took very little interest in the technical development; all he cared about was getting some useful results at the end of it. But as a result of this willingness to let people play, very marked contributions to the development of nuclear technique were made by Wynn-Williams, Lewis and others.

In a rather similar detached spirit Rutherford tolerated the other kinds of research work in the laboratory. He was never really interested in subjects outside nuclear physics, but he recognized they had their place in the sun provided that place was small compared with that occupied by nuclear physics.

He was at his best in presiding over the fortnightly meeting of the Cavendish Physical Society. At the first meeting of the year he gave an enthusiastic review of the previous year's work of the Cavendish. Thereafter he would preside over the meetings, and however involved the presentation was, Rutherford would emerge, often from apparent slumber, with some pertinent question or with a masterly summing up.

Other characteristics of Rutherford's ways of life are well known. He was a great reader with wide tastes. On Sundays he was always to be found on the Gog Magog Hills, usually with Fowler, Aston and Taylor ; figure 15 shows this foursome in South Africa. On Sunday evenings he dined in Trinity, very often interchanging indiscretions with Fowler and others in the booming voice for which he was famous.

§6. THE NEW CAVENDISH LABORATORY

One of Rutherford's last duties in Cambridge was to prepare for the new building of the Cavendish. He took no particular pleasure in this since, as he often said, it was building for his successor. But he answered the request of the University to raise an endowment fund with characteristic ability. He called on Eddington to write his brilliant pamphlet on the Cavendish Laboratory; he gave instructions for an official document to be prepared by myself; and he brought in the Chancellor to look for big money. Having obtained the great Austin benefaction, with characteristic promptness he sent his boys round the Continent at his expense to look at the best laboratories, and when they came back he gave them a pretty free hand in designing the laboratory. He took some pleasure in spending a little of the money on two high-voltage sets and on a cyclotron. " There won't ", he said, jingling his money, " be very much left for my successor ". Figure 16 shows the million-volt apparatus in the high-voltage laboratory, and figure 17 the cyclotron.

§7. CONCLUSION

Although Rutherford died in 1937, his influence was a major factor in the scientific supremacy of Britain in the war. The Senior Staff and research students of the laboratory, together with members of other physics schools, were mobilized in the first days of the war and developed for Britain and the allied cause the centimetric radar which turned the tide of the U-boat battle, directed the bombing of Germany and helped decisively to sink the Japanese fleet. The Liverpool branch of the Rutherford School, with Frisch and Peierls from Birmingham, initiated the work on the atomic bomb which ended the war with Japan. One can only wish that Rutherford had been alive to deal with its consequences.

THE EXPERIMENTAL BASIS OF ELECTROMAGNETISM: THE DIRECT-CURRENT CIRCUIT

By NORMAN R. CAMPBELL and L. HARTSHORN

MS. received 25 *March* 1946

ABSTRACT. The purpose of the enquiry, of which this paper forms the first part, is to show to what extent the working principles of electromagnetism can be soundly based on real experimental facts—that is to say, on experiments that have actually been performed, as distinct from the imaginary experiments, which are common to most expositions of the subject, but which are either quite impracticable or incapable of being performed with an accuracy that would be regarded as significant to-day. By maintaining this sharp distinction between fact and theory we aim at removing from the subject much of the confusion which, as the literature shows, is a continual source of trouble.

This section begins with an outline of the general principles of measurement. Current, resistance, conductance and voltage are then established independently of a knowledge of any other magnitudes. Ohm's Law and the conception of e.m.f. follow. An examination of the relations between these magnitudes and the geometrical and mechanical magnitudes then leads to theoretical conceptions like potential, and to a consideration of the status of Ampère's two laws. That for the force between two circuits is well established and defines the unit of current. That for the torque on a compass needle near a circuit is less well established ; but it leads to the recognition of *H* in a non-magnetic medium as a defined magnitude, whose significance is found to be independent of the existence of magnets.

§1. INTRODUCTORY

A REMARKABLE feature of the literature of the last decade is the number of papers that have been published on the fundamental magnitudes of classical physics. Even more remarkable is the divergence of opinion expressed on such well worn themes as the precise significance of the magnitudes of everyday electromagnetism. The experimentalist must occasionally ask himself whether ideas that have been successfully applied in an ever-widening field for so many years can really rest on such uncertain foundations.

There seems to be fairly general agreement that the working concepts are defined by the processes adopted for their measurement, but there is no sort of agreement as to what is the essential character of these processes. Some magnitudes appear to be associated with many processes, and there is no agreement as to their relative status.

Many difficulties have arisen from the fact that the basis of the subject has almost invariably been sought in processes that are not really practicable—that is to say, in theoretical experiments rather than in experimental facts. Experiments with point charges and magnet poles are well known examples. Every experimentalist soon learns that, whatever part such experiments may have played in the history of the subject, the best approximation he could make to them would to-day be considered quite worthless as evidence in support of his working principles. He realizes also that disputes about such experiments are almost meaningless in relation to science as practised in the laboratory. The very fact that all the doubt and discussion about these theoretical ideas have had no discernible effect on the course of experimental enquiry is sufficient to show that there is no very close connection between the two. The world-wide agreement among physicists in practical matters shows that we must be working on some common basis. It is curious that it should be so difficult to express it in terms on which we are all agreed. The present authors are of the opinion that the difficulty can be overcome by considering those processes only that are actually employed in practice; and in this paper we propose to show that all the working laws of electromagnetism can be soundly based on experimental operations that will be generally recognized as forming part of present-day technique. Every physicist will agree that the basis of the subject is to be found in experimental facts. If, therefore, we can derive our working principles from experiments that are frequently performed in the normal course of physical enquiry, or that are at least representative of existing technique, we shall have established our basis clear of theoretical ambiguities.

Our enquiry shows that the basic experimental laws are not those which are given most prominence in the customary expositions of the subject. It is therefore the more desirable that their importance should be generally recognized, and particularly so in the work of establishing units and standards. The work of the various international conferences has been of great practical value in securing international uniformity, and it is of the utmost importance that it shall not become confused with irrelevant theoretical considerations. The international committees are primarily concerned with securing agreement on the basic facts: their decisions will only have their proper effect on scientific practice so long as the relation between these facts and our working principles is clearly understood.

§ 2. GENERAL PRINCIPLES

By saying that the laws are to be soundly based on the experiments we mean that the development must follow an order that may be described as logical, in the broad sense in which that term is normally used in scientific investigations: we must not assume the truth of any proposition until it has been established. Thus, we must not assume that we can establish any algebraic law relating different properties until we have shown that these properties are measurable by processes that do not depend on the truth of that law.

It follows that we cannot begin with any algebraic law. The subject necessarily starts with qualitative observations, which first become quantitative when we have devised an operation whereby we can make consistent judgments of equality in respect of some observed property, and another operation which possesses the characteristics of addition. We can then, by successive addition, build up a standard scale for that property, and make measurements by establishing judgments of equality between the things measured and determinate parts of the standard scale. The measurement of length is a very familiar example of processes of this type, but it is not generally realized that the basic measurements of electrical properties are of the same kind.

Having established in this way the independent measurement of at least two properties, we are in a position to establish algebraic laws relating them—that is to say, relations between our observations of these two quantities that can be represented by algebraic equations. Such independent observations can never, except by the rarest fluke, show a relation of equality: they may, however, prove proportionality. It follows that the simplest algebraic law that we can establish will take the form $y=kx$, where x and y represent the two magnitudes that are independently measurable and k is a constant of proportionality which can be determined for any given system of the class to which the law applies from the measured values of x and y. We may now find that laws of the same form apply to other classes of similar systems, but that the constant k varies from one class to another, and is characteristic of some recognizable feature of the class. Thus k becomes established as a measure of this new property, which may be described as measurable by application of the law $y=kx$ to direct measurements of x and y. Such processes of derived measurement enable us to establish as magnitudes properties that are not additive, such as density and resistivity, and that may be described as *qualities*, to distinguish them from the additive properties or *quantities*. More complicated algebraic laws may take the form

$$y=kf(x_1, x_2, \dots x_n),$$

where $f(x_1, x_2, \dots x_n)$ denotes some definite function of any magnitudes $x_1 \dots x_n$ that have already been established by either independent or derived measurement. We may regard $f(x_1, x_2 \dots x_n)$ as a new magnitude z, which we shall call a *defined magnitude*. From an experimental point of view the law reduces to one of proportionality as before.

The value of the constant k in the above laws will obviously depend on the choice of units for x and y, as well as on the property which has been recognized as a derived magnitude. We may symbolize this fact by writing the equation in the form

$$y=S.\ a.\ f(x_1, x_2 \dots x_n),$$

where a denotes the derived magnitude and S is the *scale factor*, which varies with the choice of units only. Sometimes the constant k turns out to be the same for every system to which the law is applicable. There is then no derived magnitude, for we always use the term " magnitude " to mean something that varies from one measurable system to another. In that case there is no need for our immediate purpose to split k into two factors; it can be regarded simply as a scale factor S; and if we decide to adjust our units of the xs and ys so that this S becomes unity, we can simplify the equation to the form $y=f(x)$. Even when k varies with the system, and a derived magnitude has to be recognized, we can always reduce our equation to this simple form for one particular class of systems by adjusting the units of the xs and ys so that $S.a$ becomes unity for this class. In other words, we adjust the units of the xs and ys until S becomes unity, and define the unit of a as the property possessed by that particular class. Examples will appear later.

We shall follow the customary practice of suppressing scale factors as far as we can consistently do so. If we have established m magnitudes, and n experimental relations between them, then by suitably choosing the units of the m magnitudes we can suppress m scale factors. If $n>m$, the remaining $(n-m)$ scale factors must appear as numerical factors in the equations. Whether derived magnitudes appear or not will depend only on whether such magnitudes have been established by the experiments.

However, it is sometimes desirable to introduce another symbol as well as the scale factor S, even when no derived magnitude has been experimentally established. For example, in a few instances the constancy of k has been established over a range of experiments so wide that it has acquired a special theoretical significance as a ***universal constant***. It is well to guard against the suppression of such a constant by separating it from the scale factor. Again, we may have theoretical reasons for thinking that k may be variable, even though we have not succeeded in obtaining conclusive experimental evidence of significant variation. In such a case we may introduce a factor besides the scale factor so as to leave open the question whether k is actually variable. Obviously we are at liberty to introduce as many factors as we find to be significant in either theory or experiment, so long as we bear in mind the limitations to the significance of each.

There remains for consideration the order in which the various magnitudes should be introduced, and in this connection we note that our algebraic laws must always remain indefinite in their application, to an extent that is represented by the experimental errors of the observations on which they are based. Moreover, it is a well known fact that when the measurement of any property depends on the measurements of others, the experimental error of the dependent measurement is always greater than that of any of the ones on which it depends. A process of measurement that has proved in practice to be significant with high precision cannot therefore be soundly based on a law that has only been established with inferior accuracy. For example, it is absurd to suppose that dielectric constant, a magnitude that has proved to be significant in a very large number of physicochemical investigations with a precision better than 0·1%, is based on a law for the forces between small charged bodies which has never been established with anything like this accuracy; and it is even more absurd to suppose that voltage, a magnitude that is measured as a matter of routine by electric power companies with an accuracy of 0·01%, and in standardizing laboratories with a precision of 1 part per million, is to be defined in terms of the work done in moving an electric charge from one point to another, an operation about which our knowledge is incomparably more vague. We must therefore recognize that the familiar definitions, common to most expositions of electromagnetism, are the premises from which the mathematical reasoning proceeds, rather than definitions of the magnitudes with which we are concerned in experimental science; and in attempting to build up the subject from its experimental foundations we must begin with the independent measurement of the magnitudes that have been established with the least uncertainty, and then pass on to derived measurements based on them, always giving priority, in any one branch of the subject, to the experiments that are known to be capable of the highest accuracy.

This means that our first stage must include independent measurements of resistance; for there can be no doubt that measurements of resistance can be made with a higher accuracy, and over a greater range, than is attainable for any other property that is recognized as being distinctly electrical in character; and, moreover, the most precise measurements of resistance are in fact independent measurements. Qualitative observations of current are, however, a necessary preliminary; and as the independent measurement of current and voltage is on the same footing as that of resistance, it will be convenient to regard these three magnitudes together as comprising the first stage, and to begin with current and the idea of the electric circuit.

We shall take for granted the existence of all the instruments commonly employed in physical laboratories. From our standpoint they are merely common objects of the modern world, just as easily identified, and far more common, than the traditional amber, catskin, and lodestone, which, indeed, have proved, on closer investigation, to be no less complicated in their structure; and, what is more to our present purpose, far less amenable to precise observation.

§3. CURRENT: A MEASURABLE PROPERTY

Construct a circuit of a most general kind by connecting in series a battery, a variable resistor and, say, a filament lamp, a neon lamp, a water voltameter, and indicating instruments of all the common types—moving coil, moving magnet, thermal, electrostatic (with shunt resistor), and electro-dynamometer or current balance.

Varying the resistor, we observe that the indications of all these devices increase and decrease together; there is something characteristic of the whole collection, and not merely of each element of it. By connecting the instruments with different materials we discover the existence of conductors and insulators, and learn that the characteristic "something" is associated with a definite closed path or circuit that must be conducting throughout. Further observation reveals the following facts:—

To any definite indication of one instrument, there are corresponding definite indications of the others, and these corresponding indications are independent of the order in which the various instruments are connected in the circuit, provided that each instrument is always connected in the circuit in the same direction with reference to the rest of the circuit. The reversal of some instruments gives rise to a change in their indications: others are unaffected by reversal. We shall describe instruments of the first class as *direction-sensitive*, and those of the second as *reversible*.

These observations suggest an analogy with the flow of water, the battery being analogous to a pump, the resistor to a throttle, the closed conducting path to a pipe, and the indicators to rate-of-flow meters of various kinds. The analogy further suggests that if we construct more complex circuits, in which several branches lead into a main conductor, the characteristic feature "current" of the main circuit may be the sum of the corresponding features of the branches:

in other words it suggests that there is a property "current", which is characteristic of electric circuits, and obeys a law of addition in branched circuits. We now establish this law and thereby establish current as a measurable property or magnitude.

We construct two circuits X and Y (figure 1) with a common limb, in which we connect our indicating instruments AA. For the purpose of simplifying the description it will be convenient to imagine that all the instruments possess pointers moving over scales of the ordinary type, which are, however, initially blank, but it will be readily understood that what we say about numbering the scales is equally applicable to scales of any type, such as a set of " weights " for the current balance, which we can mark with the appropriate numbers in due course, and so on. In the branches B and U we have also indicating instruments. U is the instrument we choose to " define our unit ": in other words, we assign the numeral 1 to some definite condition of this instrument. For this country the legal standard is a current balance, and the definite condition is one of equilibrium with a certain standard weight on one pan of the balance. B, like AA, has a blank scale. We first break the circuit Y, adjust R_X until U is in the standard condition, and then mark all the scales of AA with the numeral 1 at the points corresponding to their indications. We then open X and close Y, adjust R_Y until AA again indicate 1, and mark the corresponding indication of B also 1. Next close both X and Y, adjust R_X and R_Y until both U and B indicate 1, and mark the corresponding indication of AA with the numeral 2. Proceeding in the same way, we use the circuit Y to mark the corresponding indication of B with the numeral 2; then with B at 2 and U in the standard condition to mark AA at 3, and so on. Having marked all the instruments AA in this way, we can insert some in the place of U and others at B, and verify that whatever the indication of the three groups of instruments they always satisfy the law of addition. Further applications of the same principle enable us to subdivide the scales to any desired extent within the limits set by the sensitivity of the instruments, and then to verify the law with high precision for a very wide range of current. We have then established the indication of such an instrument as a unique and precise numerical measure of the current in any branch of a circuit in which it is included.

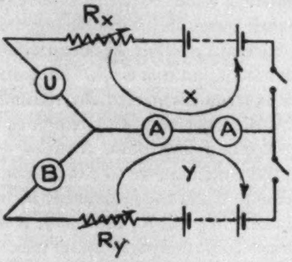


Figure 1.

It is important to notice that the measurement of current in this way is completely independent of the law of operation of the indicating devices. We may notice that many direction-sensitive instruments have scales that lead to the recognition of some " linear law "; and that the scales of many reversible instruments suggest a " square law " ; but our ideas of the magnitude " current " are not founded on any such law. The highest precision will be obtained when the indicating devices consist of the combination, well known in all standardizing laboratories, of a shunt resistor and a potentiometer, the indication being the setting of the potentiometer " slider " which corresponds with a state of balance of the potentiometer, but we must suppose that the scale traversed by the slider is initially blank, and that the combination is calibrated as an ammeter by the process we have described.

§4. CURRENT: A DIRECTED MAGNITUDE

The fact that the pointers of some instruments move in the opposite direction when the connections to their terminals are interchanged does not itself introduce any question of negative current; in effect we obtain a different instrument by the reversal, and must re-calibrate it, but that is all. Each direction-sensitive instrument must therefore have its terminals marked, say, red and black, and must always be connected in circuit in the same way with reference to the battery, so that its indications always correspond with those obtained during the calibration. If, now, bearing in mind this condition, we take the circuit of figure 1, and reverse both the battery in the circuit Y and the instrument B, we find, on closing both circuits X and Y, and adjusting R_X and R_Y until both B and U indicate the same current, that A then indicates zero current, whatever may be the current indicated by B and U. If our law of addition is to be generally applicable to branched circuits, we must therefore recognize that the sum of two equal currents may be zero; in other words, that there are such things as negative currents.

We adopt a simple convention to enable us to allocate signs to currents. We use direction-sensitive instruments, and mark their terminals so that during calibration the red terminal of one instrument is always connected to the black one of the next in the same circuit. Then at any junction the currents indicated by instruments with their red terminals connected to the common point are given the + sign, and those of instruments with the black terminals connected to the common point the − sign.

We now establish as an experimental fact Kirchhoff's First Law. The sum of the currents in all the branches meeting at any junction is always zero. This is one of our basic laws—grounded directly on observed facts and independent of any theory of electricity.

§5. RESISTANCE

The idea of resistance as a magnitude characteristic of the various parts of an electric circuit first arises when we observe that the insertion of a resistor in, say, the simple circuit first considered diminishes the indications of all the instruments, and that the insertion of a second in series with the first decreases it still further. We are led to define resistors as equal in resistance when the substitution of one for another causes no change in the indicator, and to establish the operation of connecting resistors in series as addition. It is scarcely necessary to enter into details. Every experimentalist knows that the most accurate measurements of resistance are always made by " simple substitution ", which means by making the above-mentioned judgments of equality; and in standardizing institutions, the standard scale of resistance is always established by constructing resistance "build-ups"—that is to say, groups of resistors connected in series in such a way that it can be shown by experiment that they satisfy the law of addition with high precision. Resistance, then, is a property defined by these operations, and measured independently of any other property; for it may be noted that only qualitative observations of the current indicators are necessary. The various bridges and balancing devices employed for the work are to be regarded as detectors rather than measuring instruments—that is to say, they enable us to make precise judgments of equality of resistance much as a comparator enables us to make precise judgment of equality of lengths. Over a wide range of conditions, resistance proves to be independent of both the circuit employed for its measurement and the property *current*.

§6. CONDUCTANCE

It will be convenient to consider also at this point *conductance*, another magnitude independently measurable, but closely associated with resistance; for the test of equality for the two quantities is exactly the same. They differ only in the law of addition, conductance being added by parallel connection instead of series. We find by experiment that for any resistor, conductance and resistance are inversely proportional to one another. The constant of proportionality is merely a scale factor which for convenience is suppressed by defining the unit of resistance and the unit of conductance by means of the same standard resistor. Conductance and resistance are then reciprocals.

§7. VOLTAGE

At first sight it would appear that we can establish a property " e.m.f." characteristic of batteries by a process almost identical with that established for resistance and equally fundamental, involving only addition by series connection and judgments of equality based on observations of no change of current on substitution in any circuit. Since the reversal of a battery produces changes of " current ", we must distinguish the two terminals as, say, red and black, and when we find that the series connection of a battery in one direction increases " current ", and in the other direction decreases it, we are prepared to recognize e.m.f. as a directed magnitude, the reversal of the battery changing the sign. Detailed experiments show, however, that the method cannot be established. The judgments of equality are found to depend on the circuit employed as well as on the batteries under comparison; and with any one circuit, the combination (3–2) for example, is never found to be precisely equivalent to the combination (2–1). Subsequently we explain the failure of the method by recognizing another property of batteries, " internal resistance ", but at this stage we can only conclude that the fundamental property *voltage* must be sought elsewhere.

We find that we have two distinct classes of electrical indicating instruments. In addition to those that we have used in order to establish the magnitude *current* (the ammeters), we have others which, unlike the ammeters, cannot form part of any circuit, since on inserting them into any circuit they invariably give rise to zero current, but which nevertheless give definite indications when connected across any component of a circuit. We shall now show that just as the instruments of the first class enabled us to establish the magnitude current in virtue of a law of addition, those of the second class (the voltmeters) enable us to establish voltage by a similar law. As in the first case, the process is quite independent of the law of operation of the instrument itself, but we may note that our voltmeters include electrostatic instruments, thermionic instruments of the electrometer type, and potentiometers. Some of these contain batteries and are found to be direction-sensitive: others are reversible and obey a different law, but all serve to establish the same magnitude. The potentiometer is by far the most accurate of the voltmeters. The fact that the same mechanism also formed part of one of the ammeters need cause no confusion: we must again suppose that the instrument has initially a blank scale, which by appropriate marking and numbering will be established as a scale of voltage.

We observe that if any voltmeter is connected in succession to points XY, YZ, XZ (figure 2), Y being between X and Z and no battery being between X and Z, then it gives a greater reading in the position XZ than in either of the positions XY or YZ. This observation suggests that a law of addition can be established for the property indicated, voltage being additive for circuit elements connected in series. The experimental procedure is closely analogous to the 3-instrument method used for current, and will be sufficiently obvious from figure 2. One voltmeter U is arbitrarily chosen as standard, and its indication when connected to the points XY, say, is marked 1. The indication of a second instrument A, also connected to XY, is also marked 1, and it is then transferred to B, and the point Z adjusted until it again reads 1. At this setting the indication of the third instrument C is marked 2. Proceeding in this way we establish our standard scale of voltage by successive additions of 1, and verify the laws of addition by proving that the indications of the three instruments always satisfy the relation $C = A + B$.

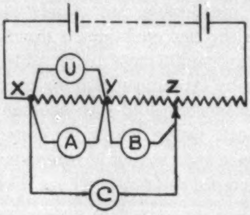


Figure 2.

In order that the law of addition may apply perfectly generally to all circuit-elements in series connection we must recognize voltages of both signs. Thus in the circuits of figures 3 and 4 a voltmeter connected to the terminals B and D may indicate zero, even when it gives finite indications in connection to BC and CD.

The sum of the voltages of BC and CD is by the law of addition zero, and, therefore, these two voltages must be equal but of opposite sign. Positive and negative voltages can be recognized by the use of direction-sensitive voltmeters with marked terminals, say one red and one black, as for ammeters. With such an instrument we find that connection to the terminals BC and CD in the case just mentioned gives the same indication provided that the same terminal of the voltmeter is connected to B of the pair BC and to D of the pair CD. Thus, reversal of the order of the voltmeter terminals when passing along a circuit indicates a change of sign of voltage. We may note here that the terminals of any component of a circuit are essentially the devices by means of which a voltmeter can be connected to two definite points on that circuit-element.

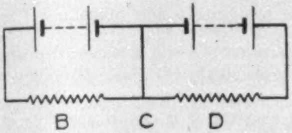


Figure 3.

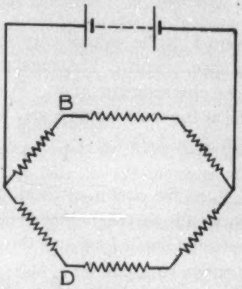


Figure 4.

We are now in a position to establish the following experimental law concerning voltage. The algebraic sum of the voltages measured between all the successive pairs of any set of terminals on any closed circuit is always zero. This law is analogous to Kirchhoff's First Law, and is of the same basic importance.

§ 8. OHM'S LAW

Having now completed our first stage and established the independent measurement of current, voltage and resistance by processes that will give, in favourable circumstances, an accuracy of the order of 1 part per million, we can now proceed to the development of algebraic laws based on these measurements. The most obvious one is Ohm's Law: for a large class of circuit-elements the terminal voltage V is proportional to the product of the resistance R and the current I characteristic of that element:

$$V = kIR. \qquad \ldots\ldots(1)$$

The proportionality constant k proves to be independent of the size and shape of the circuit element and of its composition; metals, non-metals, electrolytic solutions, etc., all have the same value if the law holds. The constant is therefore no more than a scale factor, and is suppressed by adjusting the units of V, I and R so that $k = 1$. Incidentally the fixing of the sign of k determines the relation between the red and black terminals of ammeters and those of voltmeters, which have hitherto been unrelated.

§9. ELECTROMOTIVE FORCE

We now find that there is an important class of circuit elements, including batteries, dynamos, thermocouples, etc., which do not obey Ohm's Law, but are characterized by a finite terminal voltage even when they are " on open circuit " and, therefore, have zero current. When these devices are included in a circuit we find that their terminal voltage varies with the current, and we establish the experimental law

$$V_0 - V = kI,$$

where V is the voltage corresponding with the current I, and V_0 is the voltage on open circuit. The constant k proves to vary with the circuit element and to be characteristic of it. We must therefore recognize a new magnitude r as well as the scale factor S and write

$$V_0 - V = SrI \quad \text{or} \quad V = V_0 - SrI. \qquad \ldots\ldots(2)$$

If, then, the circuit external to the generator consists of n circuit elements in series, all obeying Ohm's Law and of resistance $R_1, R_2 \ldots R_n$, we have, by applying Ohm's Law and the law for the summation of the voltages of a closed circuit,

$$IR_1 + IR_2 + \ldots IR_n - (V_0 - SrI) = 0.$$

Note that a negative sign has been attached to the expression for the voltage of the generator in this equation. This is because experiment shows that the order of the voltmeter terminals must be reversed when passing from the resistor elements $R_1, R_2, \ldots R_n$ to the generator. When we examine this relation we perceive that a great simplification can be achieved by choosing the unit of r so that the scale factor S becomes unity, in which case the equation becomes

$$I\left[\sum_1^n R + r\right] = V_0. \qquad \ldots\ldots(3)$$

It is natural therefore to interpret r as a resistance inherent in the generator, and with this interpretation we obtain Kirchhoff's Second Law in the familiar form

$$\Sigma IR = E, \qquad \ldots\ldots(4)$$

the summation now including every element of the circuit, and E, the e.m.f. of the generator, being defined as its *open-circuit voltage*. The introduction of this new defined magnitude is not really necessary, but it has proved to be useful in practice. We can establish that the magnitude so defined is additive in series connection and capable of either sign, so that the general expression for Kirchhoff's Second Law becomes

$$\Sigma IR = \Sigma E.$$

The unit of e.m.f. is, of course, the same as that of voltage, and the well known Weston standard cell can be regarded as a standard of either e.m.f. or voltage, though voltage is unquestionably the more fundamental property. Another point to notice is that we have extended the notion of resistance; for the property r is not measurable by the process by which the magnitude *resistance* is primarily defined. Experimental justification for identifying the two properties can be obtained later by showing that correlation can be established between them by means of measurements with alternating current.

§10. ROUTINE METHODS

Having now established the basic laws of direct-current networks we can use them in order to establish all the bridge networks and potentiometer circuits and the great variety of operations by means of which measurements of current, resistance, voltage and e.m.f. are made in everyday practice. We can now use Ohm's Law for example for the purpose of measuring voltage in terms of resistance and current when that procedure proves to be more convenient than the one described. It should be noted here that when we say that two different processes serve to measure ***the same magnitude***, we mean that experiment has shown that the two processes give the same result apart from a scale factor.

Routine methods differ very considerably from those we have described, mainly because the operations needed in order to establish a law are not necessarily those most convenient for subsequent applications of that law. When once our basis has been firmly established, our procedure is mainly dictated by consideration of economy of labour and the consistency of our measurements from day to day and year to year. The ease of construction of the required apparatus and its permanence become of dominating importance. In practice we find that a standard series of resistors is the most easily constructed, most permanent, and most complete as regards range and fineness of subdivision of all electrical standards. The general tendency is, therefore, to base most operations on measurements of resistance. Thus the operations we have described may appear at first sight as unfamiliar and not representative of actual practice. The practical application of Kirchhoff's First Law may take the form of a search for " leakage " or a " stray current " in order to explain some discrepancy in the observations. The experiments we have described are admittedly not frequently made deliberately with the idea of establishing a law. We think, however, that the experimentalist will recognize them as the logical equivalent of the mass of indirect experimental evidence for the truth of these laws which he has acquired in the course of his work.

We have now reached a stage at which it becomes necessary to establish relations between the basic electrical magnitudes and those of other branches of physics. Probably every experimentalist will be prepared to take for granted length, mass and time as three independently measurable magnitudes. We find, moreover, that over long periods of time our measurements of these three quantities show greater accuracy and consistency than those of any other physical quantities, and we may well therefore regard them as fundamental. We shall here also take for granted other non-electrical magnitudes that are usually held to be established in terms of length, mass and time by well known relations. The means by which these relations can be established is a separate subject.

§11. FARADAY'S LAWS OF ELECTROLYSIS

Among the simplest of the laws that we can now establish are Faraday's Law for the various voltameters and electrolytic cells. We find that if I denotes the current in the cell, the mass, W, of any element liberated in time t can be represented by the relation

$$W = \beta . I . t. \qquad ......(5)$$

The constant of proportionality β is found to be characteristic of the element, and is therefore established as a derived magnitude, measurable for the various elements in terms of mass, current and time. We may fix the unit of β by making the scale-factor unity correspond with existing units of W, I and t, as when we measure β in " grams per ampère-second ". Alternatively we may fix the unit of current by assigning a definite value to β for silver (which is the element for which the most accurate experiments have been made), when the scale factor is unity and the units of W and t are those already chosen. The existing " International ampère " was defined in this way at a time when the silver voltameter was considered to be the best standard for current measurements, being easily available at all times and in most places, and giving very consistent results within a prescribed range of conditions.

§ 12. POWER DISSIPATION

When we study the filament lamp and the neon lamp of our original circuit we find no simple law relating either the luminous flux or the heat developed to the current or voltage alone, but calorimetric measurements show that heat is produced by the lamps at a rate proportional to the product of their current and terminal voltage. Further experiments with other circuit components show that this law is quite general, and that the proportionality constant is no more than a scale factor. The units are therefore chosen so that we may express the relation in the simple form

$$P = VI. \qquad ......(6)$$

By the principle of the conservation of energy the dissipation of energy in a circuit must be accompanied by a uni-directional change in some part of the system, which thereby loses that energy. In our circuit the change is found to be chemical and to occur in the battery. We may say that the battery loses chemical energy equivalent to the thermal energy gained by the rest of the circuit, and that the difference between gain and loss of energy corresponds with the difference in the sign of V and, therefore, of VI for the two parts of the circuit.

§ 13. THE LAWS OF RESISTORS AND CONDUCTORS

Any object that may form part of an electric circuit, so that it may be characterized by a finite current, may be called a *conductor*, but we do not find it possible to determine a definite conductance and resistance for every conductor. Those for which definite values may be found obey Ohm's Law, and may be called *resistors*; but others, like the neon lamp, do not obey Ohm's Law even approximately: they are often called non-linear conductors, since they are characterized by a non-linear relation between current and voltage. It may be remarked that the law of power dissipation holds for all conductors, both linear and non-linear.

As soon as the measurement of resistance and conductance has been established, it becomes possible to establish laws relating these properties to temperature.

The experiments are complicated by the dissipation of power in the resistor, which restricts the range of current that can be employed in the detector circuit, but within limits they are practicable. When Ohm's Law is applied to any such resistor, due account must be taken of the variation of its resistance with temperature, and if this is done the law is found to hold good. On the other hand, the relation between voltage and current is non-linear, except for very small currents, on account of the variation of the temperature, and therefore of the resistance, with current. The filament lamp is therefore an example of a non-linear conductor or resistor which obeys Ohm's Law, while the neon lamp is a non-linear conductor which does not obey Ohm's Law. The most precise experiments show that all resistors are non-linear if we cover the widest possible range of current and voltage, and any measured value of resistance or conductance is therefore of significance only within a limited range of current and voltage.

We can also establish relations between the resistance and conductance of conductors of given material and their size and shape, but we must first notice that we have so far assumed that our circuits and, therefore, their components are " linear " in a sense quite different from that just mentioned. We have regarded the circuit as defined by a succession of points or terminals, such that voltmeters connected across these points, and ammeters connected between them, give definite indications, characteristic of the circuit or its elements. The circuit is therefore linear in the sense that it can be represented by a line having the same value of current at every point on it, and that the location of any voltmeter terminal on the circuit is as definite as that of a fixed point on this line. We immediately think of thin wires as providing the simplest method of achieving the conditions described, and although we have mentioned neon lamps and voltameters, which are certainly not thin wires, we must suppose that they have been joined by thin wires, and that each terminal is some definite point on such a wire. We think of the wire as providing a linear path for the current at the crucial points. When we come to deal with conductors of large cross-section, we think of current as entering terminal areas rather than terminal points, and we find that, by fixing one voltmeter terminal at some definite point on such a conductor and using the other as a probe, we can trace on the surface of a conductor lines of constant voltage called, for reasons to be given later, "equipotentials". Thus a terminal is now regarded as a device by means of which a voltmeter can be connected to an equipotential rather than a point. The experiments of high precision, by means of which the relations between the electrical properties of conductors and their geometrical properties have been established, have been almost entirely confined to straight conductors of constant cross-sectional area. Experiment shows that when these constitute parts of circuits, the equipotentials lie in parallel planes, and that the following law holds good :

$$\frac{I}{V} = \sigma \frac{A}{l}, \qquad \ldots\ldots(7)$$

where I denotes the current in the conductor, V the voltage corresponding with any two equipotentials separated by a distance l, A the cross-sectional area of the conductor, and σ is a constant for any one material, but varies with the material. For completeness we should include a scale factor, but since I/V, by Ohm's Law, measures a conductance, the unit of which is fixed by other considerations, and the units of A and l are also fixed by other considerations, it is usual to make the scale factor unity for this particular set of units, thereby fixing the unit of the derived magnitude σ, conductivity.

By inverting the equation (7) we can show that the same experiments also establish a second derived magnitude, resistivity ρ, which becomes the reciprocal of σ, if the same convention is followed in fixing the unit. We have

$$-\frac{V}{I}=R=\rho\frac{l}{A}. \qquad \ldots\ldots(8)$$

It is obviously desirable that we should be able to deal with conductors of other shapes, and we therefore try to establish laws of greater generality. At this stage we find it necessary to introduce purely theoretical conceptions. We regard the conductor as the field of a vector $\mathscr{E}$ which is equal in magnitude but opposite in direction to the gradient of a scalar function v, the potential. It follows that the line integral of this vector between any two points is independent of the path followed, and simply equal to the difference of v for the two terminal points, and, moreover, for any closed path the line integral is zero. The function clearly has properties akin to those of voltage, and we can regard it as a generalization of voltage if it satisfies the condition that the difference of potential for any two points between which a voltage can be measured is equal to that voltage. As a generalization of current we introduce another vector $\mathscr{I}$, the "current density", which satisfies the two conditions: (*a*) that its flux across any terminal area for which a value of current can be measured is equal to that current, and (*b*) the "law of continuity",

$$\mathrm{Div}\ \mathscr{I}=0, \qquad \ldots\ldots(9)$$

which implies that the flux of $\mathscr{I}$ across any closed surface is always zero, and which we may regard as a generalization of the law that the current is the same along any linear circuit. We now write as a generalization of the law established for straight conductors of uniform cross-section,

$$\mathscr{I}=\sigma\frac{dv}{dn}=\frac{1}{\rho}\frac{dv}{dn}, \qquad \ldots\ldots(10)$$

where n is the normal to an equipotential surface, σ is the conductivity, and ρ the resistivity of the medium at the relevant point. This relation implies that the lines of flow of $\mathscr{I}$ are everywhere normal to the equipotential surfaces and therefore identical with the lines of flow of $\mathscr{E}$. These theoretical relations are postulates, for the truth of which we can never obtain direct experimental evidence; for $\mathscr{I}$ and v, and even $\mathscr{I}\,ds$, the current crossing a small area ds, are not measurable: it is always impossible to insert an ammeter into the conducting medium without altering the conditions appreciably. The postulated relations have, however, been found to lead to all the relations that have been established experimentally for conductors of various shapes. As we have seen already, equipotential lines can be traced experimentally on the surfaces of solid conductors, and in liquid conductors, equipotential surfaces having given voltages from one electrode may be traced out in a similar way. In electrolytes, moreover, it can be shown that the thickness of deposited metal has some relation to current density. The postulated quantities are therefore not entirely without experimental significance, though the experiments in which they are significant are only of low accuracy. We shall later find applications for these conceptions in other branches of the subject.

§ 14. AMPÈRE'S LAW

We have seen that the basic laws in any branch of the subject are those which have been established by, or which govern the operation of, the instruments of highest precision available for work of that kind. It follows that the fundamental law of electrodynamics is the law of operation of the current balance or electrodynamometer, which is, broadly speaking, Ampère's Law, although we shall find it desirable to express it in terms different from those actually employed by Ampère.

Very simple experiments are sufficient to show that the mechanical force between two circuits is proportional to the product of the currents they carry; but the relation between the constant of proportionality and the geometrical properties of the circuits is so complicated that its discovery by Ampère was described by Clerk Maxwell as one of the greatest achievements in physical science. Ampère, by making inspired guesses based on the examination of circuits of simple form, found the form of the constant almost as soon as he discovered the proportionality. Subsequent work has served only to confirm the law with higher and higher precision, and to define the conditions in which it holds good. Ampère expressed his law in terms of the forces between circuit elements, and thereby introduced an ambiguity that has been the foundation of many paradoxes. But this ambiguity can be removed by translating his law into terms of the shapes and geometrical relation of complete circuits. It then becomes

$$F = kI_1 I_2(-\partial N/\partial q), \qquad \ldots\ldots(11)$$

where F is the force tending to increase a coordinate q; I_1, I_2 are the currents in the two circuits; and N is defined as

$$N = \oint_1 \oint_2 \frac{dl_1 \, . \, dl_2 \, . \cos\theta}{r}, \qquad \ldots\ldots(11.1)$$

where dl_1, dl_2 are infinitesimal elements of the circuits 1, 2; r the distance, and θ the angle, between them. Thus N and $(-\partial N/\partial q)$ can be determined by geometrical measurements on the circuits.

Equation (11) is limited by its form to strictly linear circuits; but it can be extended in an obvious manner to any pair of circuits that can be regarded as each made up of linear filaments, rigidly connected, the proportion of the circuit current passing through each filament being known. But there is another limitation, usually expressed by the statement that the circuits must be composed of, and the space between them occupied by, a medium that is non-magnetic. We are not yet in a position to define " non-magnetic ", and when we are we shall find that none of the materials usually present in the experiments to which (11) is applied are, strictly speaking, non-magnetic. However, a sufficient condition for (11) to be true can be stated definitely; it is that every medium in the neighbourhood of the circuits should be either air or equivalent to air, equivalent meaning that, if any part of the medium that is not air is replaced by air, or *vice versa*, no appreciable change in the force between the circuits is produced. Equivalence thus depends on what changes are appreciable—that is to say, on the sensitivity of the measurement of force.

In this form and with these limitations the law (11) has been confirmed by many investigations with an accuracy that has steadily increased down the years. The most recent experiments at the National Physical Laboratory and the National Bureau of Standards agree and confirm the law to about 2 parts in 100,000.

In most treatises k is written μ, and treated as a derived magnitude whose value may vary. The basis of this practice is the belief that, if the air and its equivalents used in current balances were replaced by a medium not equivalent to air, and therefore magnetic, then, so long as the medium were uniform, (11) would still be true if k were allowed to take a different value. No experiments comparable in accuracy with those in air have been made in any other medium; and it is impracticable to use in such experiments the most important magnetic materials, which are solid and would therefore hamper greatly the measurement of force. Accordingly the direct evidence that k is a derived magnitude, and that it ought to be split into two factors $S\mu$, is of a quite different nature from that for (11), subject to the condition stated above. On the other hand, there are good theoretical grounds, which will be considered later, for thinking that k may be a derived magnitude. In these circumstances it is convenient to take advantage of the liberty noted in § 2 and to write

$$F = S\mu I_1 I_2(-\partial N/\partial q), \qquad \ldots\ldots(12)$$

leaving open for the time the question whether μ is a derived magnitude. It should be noted that (12) still requires that the medium about the circuits should be uniform. If it is not uniform in the sense that an interchange of the medium in one region with that in another might produce a change of the force, a law of the form (11) or (12) can give no account of the matter; for they contain no term that depends on the position of any material that does not form part of the circuits.

The law of the current balance has now been established with such accuracy that it can be used to define the unit of current. The law is (11) for the special case in which $I_1 = I_2 = I$, and we may write it

$$F = S\mu_0 I^2(-\partial N/\partial q), \qquad \ldots\ldots(13)$$

the factor μ_0 serving to remind us that we are limited to an air-equivalent medium. The ampère has been defined by assigning to $S\mu_0$ in this law the value 10^{-2} when F is measured in dynes, or 10^{-7} when F is measured in m.k.s. units. The constant μ_0 in (13) is commonly termed the permeability of free space; it is perhaps worth noting that the relations implied by this term lie quite outside the basic facts that serve to determine the unit of current.

§ 15. MAGNETIC FIELD STRENGTH

Ampère, in his investigation of Oersted's well known discovery, also enunciated the law of the " moving magnet " indicator. Instruments of the type so named have been found to be incapable of the same accuracy as the current balance, and the law has therefore never been established with high accuracy. It has, however, been found to be consistent with the available evidence, and we shall therefore accept it as a hypothesis. It can be written in the form *

$$\boldsymbol{T} = \boldsymbol{m}SI\oint \frac{\boldsymbol{dl}\times \boldsymbol{r}_1}{r^2} = Sm I\boldsymbol{G}, \qquad \ldots\ldots(14)$$

where $\boldsymbol{T}$ represents the torque on a small compass needle, about its axis of rotation, arising from the presence in its neighbourhood of a circuit carrying a current I; S is the scale factor; $\boldsymbol{dl}$ represents in length and direction any element of the circuit; $\boldsymbol{r}_1$ is a unit vector in the direction from this element to the magnet; × represents a vector product; r is the distance between the element and the magnet; the integration is taken completely round the circuit; and $\boldsymbol{m}$ is a derived magnitude characteristic of the needle, its direction always perpendicular to the axis of rotation.

The form of the law requires that the dimensions of the magnet should be small compared with any r, in order that $\boldsymbol{G}$, the integral, should be definite. A limitation is also imposed on the nature of the medium surrounding the magnet and circuit; a sufficient, but perhaps not a necessary, condition is that the medium should be air or its equivalent in the sense defined above.

The experimental evidence for (14) is scanty and usually indirect. Thus, though the law is implied in the well known experiments with sine- and tangent-galvanometers, those experiments do not enable $\boldsymbol{m}$ to be determined in terms of current, torque and the geometrical magnitude $\boldsymbol{G}$; for the torque is measured in terms of that which the needle experiences in the absence of a current circuit, and it is assumed that this torque is also proportional to $\boldsymbol{m}$. In particular, there is very little evidence that $\boldsymbol{m}$ is independent of I and $\boldsymbol{G}$, as the law requires; indeed, there is indirect evidence that, for large values of I, $\boldsymbol{m}$ is not independent of I. However, such doubts are of little consequence, because $\boldsymbol{m}$ is not an important derived magnitude.

The real importance of (14) arises from its implication that two circuits for which $I\boldsymbol{G}$ is the same have the same effect on any compass needle. This law proves to be a special case of one more general; the torque on a compass needle is only one of several effects that can be observed in the neighbourhood of current circuits, e.g., the deflection of a cathode ray and the production of an e.m.f. in a rotating coil. All such effects (so long as the media are non-magnetic) are determined by $I\boldsymbol{G}$ in the sense that two circuits for which $I\boldsymbol{G}$ is the same produce the same effect, and that two circuits whose $I\boldsymbol{G}$s are equal and opposite compensate each other and produce no resultant effect.

The experimental evidence for this statement is much stronger than that for (14). Some of it arises from an intimate mathematical connection between $\boldsymbol{G}$ and N (it will be stated explicitly in a later section), which was used at the National Bureau of Standards in some of the experiments on the current balance. Instead of determining N wholly by geometrical measurements, they determined

---

* This form is taken from G. P. Harnwell, *The Principles of Electricity and Magnetism* (McGraw-Hill, New York and London, 1938). This book seems to be representative of the working ideas of the experimentalist of today.

an important part of it as the ratio of the $\boldsymbol{G}$s for the fixed and moving coils when these were placed suitably relative to each other; this they did by determining the ratio of the currents in the two coils at which the resultant effect on a compass needle vanished. The fact that they could thereby obtain consistent values of N for different pairs of coils, and that this method of determining N led to results agreeing with those of the National Physical Laboratory, who relied entirely on geometrical measurements, is very complete evidence for the significance attributed to $I\boldsymbol{G}$ as well as for (11).

Another part of the evidence for the significance of $I\boldsymbol{G}$ rests on the " Schuster magnetometer ", which is now recognized as the most accurate instrument for the measurement of the earth's magnetic field. (We are not yet in a position to define this quantity; all that we need to know about it here is that it is something that can be compensated by a current circuit.) This instrument consists essentially of a current circuit constructed so that the value of $\boldsymbol{G}$ for a small region fixed relative to it can be determined with high accuracy, some form of field detector being mounted within this region—in practice a small compass needle, a rotating coil, or a vibrating circuit mounted at the centre of a Helmholtz coil-system. The field within the small region is measured by observing what current must flow in the circuit in order that the detector shall indicate zero. Experiments with this instrument establish with high accuracy that the compensating current is inversely proportional to $\boldsymbol{G}$ at the relevant point.

These facts lead us to define a magnitude $\boldsymbol{H}$, characteristic of points in the neighbourhood of a linear circuit, by

$$\boldsymbol{H}=S.I.\boldsymbol{G}.=S.I.\oint\frac{\boldsymbol{dl}\times\boldsymbol{r}_1}{r^2}. \qquad \ldots\ldots(15)$$

We insert the scale factor merely to allow for changes of unit. In accordance with general practice we shall call $\boldsymbol{H}$ the " magnetic field strength due to the current circuit ". Notice that it is not a magnitude established by experimental facts, but one laid down by definition; its importance depends on the fact that the quantity so defined proves to be significant in a wide range of experiments of the kind mentioned above. The definition can be extended to a non-linear circuit, so long as it can be regarded as a collection of linear circuits, by making $\boldsymbol{H}$ (which is a vector) mean the vector sum of components each associated with one of the linear circuits. In this extended meaning $\boldsymbol{H}$ retains its significance.

Purely mathematical reasoning is sufficient to show that the vector $\boldsymbol{H}$, defined by (15), is the negative gradient of $I\omega$, where ω is the solid angle subtended by the circuit at the point to which $\boldsymbol{H}$ refers; and $I\omega$ is therefore recognized as the magnetic scalar potential. It follows that the line integral of $\boldsymbol{H}$ around any closed path that does not link a circuit is zero, and that the line integral around a closed path that links n times in the same sense a circuit carrying a current I is given by

$$\oint\boldsymbol{H}.\boldsymbol{dl}=S.4\pi nI. \qquad \ldots\ldots(16)$$

We may remark that the two units of H that are in practical use arise naturally from equations (15) and (16). The oersted, the unit of the c.g.s.μ system, was chosen so as to simplify (15) by making $S=1$. The "ampère-turn per metre", the unit of the m.k.s. system, was chosen so as to simplify (16) by making $S.4\pi=1$. There is no difference in principle between the two choices, and usage alone can show which is the more judicious.

Finally a word about magnetic media. If the medium surrounding the circuit is magnetically heterogeneous, (14) ceases to be true; circuits with the same IG are not in generally equivalent, so that H loses its significance if it is defined by (15), and there is no reason to adopt that definition. Whether in these circumstances there is any other quantity which plays in any sense the same part as H plays when the medium is non-magnetic is a question that must wait until a later section.

If the medium is uniform, but not equivalent to air, the scanty evidence available suggests that circuits with the same IG are still equivalent. Since (15) merely expresses this equivalence, there is no need to introduce a factor μ into it; whether it is desirable to do so, and what modification, if any, is required in (14), are again questions that must be postponed.

At this stage we may regard all the basic principles of circuits under direct current conditions as having been established. We have considered so far only currents, voltages and circuits that are invariable with time, and mainly materials that are non-magnetic. In subsequent parts we propose to deal with varying currents and voltages, which lead to the conceptions of inductance and capacitance, and so to the principles of magnetism, electrostatics and electromagnetic waves.

In concluding the first portion of our enquiry we would emphasize that we are not trying to establish a new system of electrical measurements, but merely to discover the logical basis of the system which is in fact universally practised. It follows that all the relations that we develop will be familiar ones, and the differences between our treatment and that of the standard treatises may appear trivial, especially in this first portion. The detail given is, however, necessary for logical completeness and to establish the method. The differences will become more significant in electrostatics and magnetism, and it is from these branches of the subject that we hope to remove many of the difficulties that frequently disturb the experimentalist. To the mathematician there are many possible ways of developing the relations of electromagnetism, and they are all exactly equivalent in logic and differ only in elegance. The experimentalist has less latitude; his practice is dictated by stubborn facts that defy representation in equations. The one course he is compelled by circumstances to follow is more important to him than the more elegant courses he might have followed in a better world. This is the excuse we offer to those who find our treatment clumsy.

LOW-ORDER MULTIPLE-BEAM INTERFEROMETRY

By S. TOLANSKY,
University of Manchester

MS. received 6 *May* 1946

ABSTRACT. A discussion is given of the various factors affecting the intensity and sharpness of the multiple-beam Fizeau fringes and fringes of equal chromatic order used for the study of surface topography. The effects of absorption in the silver film, phase condition, linear displacement of the beams, finite size of source, departure from parallelism and source line width are considered. It is shown that Fabry-Perot fringes, multiple-beam Fizeau fringes and fringes of equal chromatic order have all the same fringe shape if certain conditions are fulfilled.

§ 1. INTRODUCTION

In a group of earlier papers (Tolansky, 1944, 1945 a, b) it has been shown how multiple-beam interference methods can be applied to the study of surface topography. It is the purpose of this paper to discuss the factors affecting the definition of such fringes. The multiple-beam interference is produced by depositing highly reflecting thin silver films upon the two surfaces between which interference takes place. The fringe definition is affected mainly by the reflection coefficient, the geometrical conditions and the nature of the light. Each of these features will be considered in turn. The discussion is simplified by treating the case of interference in an air film ($\mu=1$) and by first reviewing the case of interference between plane parallel silvered surfaces (i.e. Fabry-Perot interferometer).

§ 2. INTENSITY OF THE FRINGES

For multiple-beam interference between two parallel plane silvered surfaces, distant t in air, with reflecting coefficient R, the well-known Airy summation formula shows that the resulting fringes have an intensity distribution I given by

$$I=\frac{I_{\max}}{1+F\sin^2\delta/2},$$

in which $F=4R/(1-R)^2$ and δ, the phase difference between successive beams, equals $2\pi/\lambda \,.\, 2t\cos\phi$; the symbols here have the usually accepted meanings, and $I_{\max}$, which is the intensity of the fringe maximum, equals $T^2/(1-R)^2$, where T is the fraction of light transmitted by each silver film. If there is no absorption, $I_{\max}=1$, but if a fraction A is absorbed by each film, then as $T+R+A=1$, the value $I'_{\max}$, involving absorption, is $I'_{\max}=\left(\frac{1}{1+A/T}\right)^2$, Airy's formula showing that it is desirable to have R as high as possible in order to produce the sharpest fringes. It is possible with silver to make R as high as 0·95 (in the red end of the spectrum), but such a silver film has so high an absorption that it is useless for interferometry. Indeed, the ***absorption*** is the important factor which limits the film thickness that can be tolerated, and hence the reflectivity. Strong (1940) has given the values of T and R for evaporated silver films, up to $R=0{\cdot}89$, and the curves are such that it is permissible to extrapolate these slightly to include $R=0{\cdot}90$, this value being convenient for calculation. From these data A is simply obtained, and thus the quantity $I'_{\max}$ can be evaluated. The result of this simple calculation is shown in figure 1, which gives the intensity of the fringe maximum as a fraction of the incident intensity for different R-values.

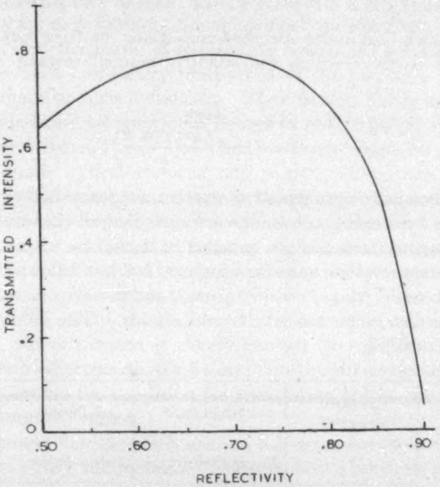


Figure 1.

This curve only obtains if particular attention is paid to purity conditions, for slight impurities seriously increase the value of A.

The fall in $I'_{\max}$ beyond a reflectivity of 0·80 is very rapid, and at $R=0{\cdot}90$ only some 5% of the incident light is transmitted. This curve shows why it is rarely possible to exceed $R=0{\cdot}87$ when using a Fabry-Perot interferometer for hyperfine structure studies. An increase of R from 0·85 to 0·90 improves the resolving power by 60%, but reduces the intensity by a factor of more than 10. Intensity considerations thus restrict the reflectivity that can be employed.

[*Note **added in proof***, 2 ***September*** 1946: Measurements recently made in this laboratory enable figure 1 to be extended to $R=0{\cdot}94$, at which point $I'_{\max}$ is only 0·75% of the incident intensity.]

Fortunately in topographical studies very intense sources are available, e.g. high-pressure mercury arcs for Fizeau fringes and powerful carbon arcs for white-light fringes. This permits of the employment of much denser silverings than are practical in hyperfine structure spectroscopy, and the loss of 95%, or even more, of the incident light can thus be tolerated, even if the fringes are observed by a microscope of × 100 magnification, which is at times the case. Reflectivities of 0·90 and more can thus be employed. With such a high reflectivity, some 70 to 80 multiple images of a lamp filament can be counted, although the tail images are weak. One can certainly assume that at least 60 beams will effectively combine in the multiple-beam interference.

From Airy's formula the half-width of the fringes (width at half intensity) as a fraction of an order is $2/\pi F^{\frac{1}{2}}$, which, for $R = 0{\cdot}90$, equals 1/30 and for $R = 0{\cdot}94$ is 1/50 of an order. Fringes have been obtained which approach the latter case.

§3. VARIETIES OF MULTIPLE-BEAM INTERFERENCE

It can be shown that under specified conditions the Airy formula holds fairly closely also for surfaces which are slightly inclined instead of parallel. By writing

$$I = \frac{I_{\max}}{1 + F\sin^2\left(\dfrac{2\pi t\cos\phi}{\lambda}\right)}$$

it is clear that identical fringe shapes can arise from separate independent variations in either ϕ, t or λ providing conditions are such that all the effective beams add up and the phase differences are constant. It will be shown later that this condition can be approached, and assuming it to hold, it follows that four distinct types of multiple-beam fringes can be formed, ***and in each case the fringe-intensity distribution is the same***, since a similar formula holds. The following table shows the fringe types possible.

Light	Constant quantity	Fringe type	Name	Equivalent filter
Monochromatic (λ constant)	t	Equal inclination	Fabry-Perot	Angular
	ϕ	Equal thickness	Fizeau	Linear
White (λ variable)	ϕ	Equal t/λ	Equal chromatic order	Wave-length
	t	Equal $\dfrac{t\cos\phi}{\lambda}$	White light Fabry-Perot	—

The last case (white-light Fabry-Perot fringes) has no application to the present problems and will be disregarded, since these fringes have already been discussed elsewhere (Tolansky, 1945 c).

The fourth column is a crude yet useful alternative manner of classifying the fringes. In effect a Fabry-Perot interferometer is an *angular* filter passing only that light incident at angles close to ϕ values for which the order of interference is integral. In a similar way, with multiple beams, the Fizeau fringes of a wedge are such that the wedge is in effect a *linear* filter (grating). Finally, with fringes of equal chromatic order (Tolansky, 1945 d) every point on the interference surfaces acts as a *wave-length* filter (more accurately, as a frequency filter. This property has been made use of for the production of a new type of interferometric colour filter, details of which will be published later).

§ 4. THE PHASE CONDITION

In the Fabry-Perot case, and in some fringes of equal chromatic order (e.g. those from a double-silvered slip of mica), t is locally constant and the emerging beams are parallel, with constant phase difference. The Airy summation is therefore obtained simply by collecting the parallel beams with a lens, since the summed effect then takes place at the principal focus.

In multiple-beam Fizeau fringes (and many cases of fringes of equal chromatic order) conditions are very different, for the fringes are effectively localized at the interference film and the paths of succeeding beams are not strictly in phase. This feature has been completely overlooked in the literature despite its key importance in obtaining high definition. Thus Benoit, Fabry and Perot (1913) employed multiple-beam wedge Fizeau fringes in their classical evaluation of the length of the metre. Adam Hilger Ltd. have used such fringes for at least 20 years for testing high-grade optical surfaces, and recently Rasmussen (1945) has described further applications, yet neither in these nor in any other case has the undoubted fundamental importance of the phase condition been recognized.

Let AB, CD be the two plane surfaces, at angle θ, then with normal incidence the paths of the 1st, 2nd and 3rd beams which meet to interfere at the point X are as shown in figure 2. It is clear that successive beams do *not* have a constant phase difference. It can be shown from the geometry that the path difference between the 1st and nth beams is closely enough $2nt-\frac{4}{3}n^3\theta^2t$. For strict Airy summation the path difference should be $2nt$; thus the phases become progressively out of step and destroy the condition for summation unless the quantity $\frac{4}{3}n^3\theta^2t$ is made small compared with λ.

Let it be assumed that the nth beam is a half-wave out of step and thus completely opposes the Airy summation. Taking this as a limit, then

$$\tfrac{4}{3}n^3\theta^2t=\lambda/2.$$

If x is the number of fringes per cm. on the wedge surface,

$$\theta=\lambda x/2,$$

giving

$$t=\frac{3}{2n^3\lambda x^2}.$$

Suppose, for $\lambda=5{\cdot}5\times10^{-5}$ cm., that n is taken to be 60, then

$$t=\frac{1}{7{\cdot}92x^2}.$$

In some instances, fringes may be 1 cm. apart ($x=1$), but more frequently they are 1 mm. apart ($x=10$), or even 0·1 mm. apart ($x=100$), the latter being the case when low-power microscopy is employed. Corresponding critical values for t are

Number of fringes per centimetre	1	10	100
Critical t (mm.)	1·26	·012	10^{-4}
d_{60} (mm.)	0·25	·025	·0025

The calculation gives only the approximate order of t for the case of $x=100$, since t is then less than a light wave and the approximations used are invalid.

An important feature has emerged; for the Airy summation to be effective the values of t should be less than the values above; in other words *the distance between the surfaces must be as small as possible indeed—usually of the order of at most a few light waves.* This is the critical fact overlooked by others, for other important features depend on it also, as shown below.

§ 5. LINEAR DISPLACEMENT OF BEAMS

The phase retardation just calculated for the successive beams is dependent upon the linear displacement of the beams along the wedge surface. Thus it is advisable to use normal incidence since this displacement increases with increasing incidence.

In the case of the study of a complex topography it is clearly essential to view interference from beams which have scanned as small an area as possible. Figure 2 shows that the higher-order beams come down from points successively further and further away from the first beam, and to a first approximation the linear separation on the surface between the 1st and nth beams is

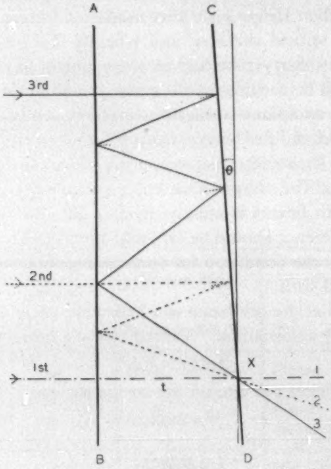


Figure 2.

$$d_n = 2n^2 t\theta = n^2 \lambda x t,$$

and substituting for t for the 60th beam (above) gives

$$d_{60} = \frac{1}{4x} \text{ mm.}$$

The values, for $x=1, 10, 100$, are included in the previous table, third row.

Again on this ground, therefore, t must be as small as possible. If t is of the order of 1/1000 mm., which is desirable, then d_{60} is of the same order, so that effectively for any feature resolvable by a low-power microscope all the beams scan within the limit of resolution.

§6. ERRORS IN COLLIMATION

Fabry (1922) first drew attention to the broadening effects in multiple-beam Fizeau fringes consequent on lack of parallelism in the incident beam arising from the finite dimensions of the " point " source. The following shows that with the values of t imposed of necessity by the phase condition, collimation defects are of no consequence.

The change in order dn produced by an angular deviation ϕ from the normal is

$$dn = \frac{2t}{\lambda} \sin^2 \frac{\phi}{2},$$

and as only small angles are under consideration

$$\phi = \sqrt{\left(\frac{\lambda dn}{t}\right)}.$$

If one is prepared to tolerate a change dn no more than 1/5 the half-width, then, for $R = 0{\cdot}90$,

$$dn = \frac{1}{150},$$

which gives the angle ϕ in degrees (for $\lambda = 5{\cdot}5 \times 10^{-5}$ cm.) as

$$\phi^\circ = \frac{3{\cdot}4 \times 10^{-2}}{t^{\frac{1}{2}}}.$$

Approximate values of ϕ for successive values of t are given below, and if a 10-cm. focal length lens is used, the corresponding diameters permitted for the source, D (in mm.), are shown in the third row.

t (mm.)	1	0·1	0·01	0·001
ϕ°	1/10	1/3	1	3
D (mm.)	0·2	0·6	2	6

Since $t > 0{\cdot}01$ mm. is excluded, a source diameter of at least 2 mm. can be tolerated, often three times this. Clearly precise collimation is only critically important for the larger t values.

§7. HIGH-DISPERSION FIZEAU FRINGES

In a recent communication (Tolansky and Wilcock, 1946) a technique of " crossed " fringes has been described in which use is made of high-dispersion Fizeau fringes. In this technique θ is made as small as possible, only the residual values due to the surface configuration being permitted. With such an arrangement the fringes are very broad and interpretation is normally difficult, but this is made simple by the " crossing " technique. The high-dispersion fringes then become a powerful method for the detection of small changes in thickness, of molecular dimensions, in accordance with the following: as θ diminishes, the Fizeau fringes broaden, and, with parallel or nearly parallel plates, this leads to a uniform tint of intensity determined by the value of t. When t is an integral number of half-waves the tint is of maximum intensity and half-way between is a minimum, the value of the latter being only about 0·25% that of the maximum (for $R=0{\cdot}90$) and thus, in effect, zero.

The fringe shape is such that the maximum tint-change sensitivity to a change in thickness occurs if a value of t is adopted such that approximately half the peak intensity is within the field of view. From the Airy formula we have the following: if δ is the change in phase from the n integral (maximum) position at which half-intensity occurs, then

$$\delta=\frac{2}{F^{\frac{1}{2}}}.$$

Supposing one can measure a change of 10% in tint intensity (a microphotometer could detect 5%), this corresponds to a phase δ' at which

$$I=0{\cdot}45\, I_{\mathrm{max}} \text{ (or } 0{\cdot}55\, I_{\mathrm{max}}\text{)}.$$

Substitution gives

$$\delta'=\frac{2\sqrt{11}}{3F^{\frac{1}{2}}}.$$

From this the fraction of an order detectable δN is

$$\delta N=\frac{\delta'-\delta}{2\pi}\simeq\frac{1}{30F^{\frac{1}{2}}},$$

giving

$$\delta N\simeq\frac{1}{570}, \text{ for } R=0{\cdot}90.$$

This corresponds to only 5 A.

Thus by selecting the correct value of t, a change of ± 5 A. in the topographical height can be measured. Such a precision and sensitivity have not yet been attained by any of the other multiple-beam methods. The method is, however, severely restricted in use and requires a microphotometer for general numerical evaluation.

§8. EFFECT OF SOURCE LINE-WIDTH

Since at normal incidence $n\lambda=2t$, the wave-length range between orders is $\Delta\lambda$, given by

$$\Delta\lambda=\frac{\lambda^2}{2t}\simeq\frac{0{\cdot}15}{t}\text{ A. for }\lambda=5{\cdot}5\times10^{-5}\text{ cm.}$$

The wave-length range for different t values is then as given below in the second row, in A.

t (mm.)	1	0·1	0·01	0·001
$\Delta\lambda$ (A.)	1·5	15	150	1500
$\delta\lambda$ (A.)	0·01	0·1	1	10

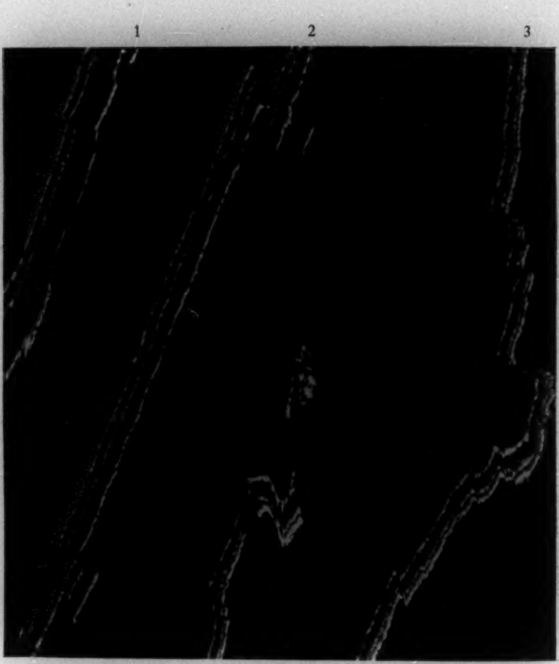


Figure 3. Fizeau fringes (mercury source) of a diamond face, area 2 sq. mm.

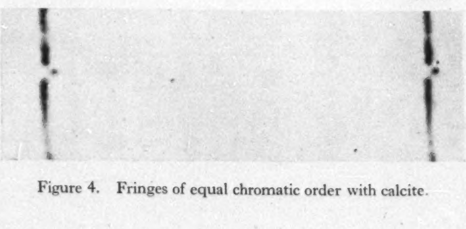


Figure 4. Fringes of equal chromatic order with calcite.

To face page 661

As in the treatment of the errors in collimation, it is assumed that 1/150 change in order is permissible. The line-width $\delta\lambda$ which can thus be tolerated is that shown in the third row. Since previous considerations limit t to be less than 0·01 mm., and preferably of the order 0·001 mm., a source line-width of 1→10 A. can be employed as compared with less than 0·01 A. needed for hyperfine structure. This fact is of the utmost importance, for this width tolerance permits the employment of an intense high-pressure mercury arc (usually, however somewhat under-run) with consequent great brilliance. The brilliance enables higher reflecting coefficients to be adopted in accordance with the intensity considerations of figure 1. Ultimately, then, much narrower fringes are obtained.

To sum up, by making t of the order of 0·001 mm., definition is greatly improved because

(1) a higher reflecting coefficient can be used;
(2) the phase condition is satisfied;
(3) linear displacement of multiple beams is reduced;
(4) errors in collimation are negligible;
(5) source line-width can be easily rendered innocuous.

As an example of the results of applying the principles discussed here, figure 3 in the accompanying plate shows the multiple-beam Fizeau fringes of 2 square mm. of a natural octahedron diamond face. 1, 2, 3 are successive orders of the green mercury line, the other fringes being the yellow doublet. The extreme sharpness is striking. The print is taken from a soft photographic plate.

§9. FRINGES OF EQUAL CHROMATIC ORDER

The former considerations as to the dependence of intensity on absorption, phase condition, linear displacement of beams and errors in collimation also apply to the white-light fringes of equal chromatic order given by a wedge. On these grounds, therefore, t should also be as small as possible for this type of fringe. Cases do arise, however, in which there is no violation of phase condition nor any appreciable linear displacement effect, e.g. doubly silvered mica film, in which locally the sides are frequently parallel, so that the conditions for the Airy summation are realized. Since the collimation defect is relatively trivial, it follows that as θ can always be reduced to a minimum by making the surfaces as near parallel as possible, then on these grounds alone fringes of equal chromatic order should be, and indeed are, somewhat sharper than the corresponding multiple-beam Fizeau fringes. In particular is this realized when higher magnifications are involved, for as already shown, in Fizeau fringes the phase condition is determined by θ^2, i.e. x^2, where x is the number of fringes per cm. This may be as much as 100, so as to give a reasonable fringe group in the field of view of the microscope. In fringes of equal chromatic order, the number and spatial distribution of fringes seen is *quite independent of the magnification used in projecting the image on to the slit;* thus θ can be made to approach zero, even for the highest of magnifications. This is an important advantage.

With a complex surface topography, there remains always a residual θ value due to the surface shape, hence the advisability of keeping t reasonably low.

However, the influence of t on fringe dispersion is much more important. Since the wave-number separation $\Delta\nu$ between fringes is $\Delta\nu = 1/2t$, and as a $\Delta\nu$ value of 1000→5000 cm^{-1} is desirable to show fine structure details with a typical Hilger medium quartz instrument, then t should be of the order 0·005→0·001 cm. Thus one arrives at a value identical with that needed to satisfy all the other conditions. Hence, again as with Fizeau fringes, t should be at most a few light-waves.

Since white light is used, line-width considerations are not involved. However, of much importance is the fact that a very intense white-light source is available in the high current-density carbon arc. This is so bright that silverings can be used which are definitely too thick even for the high-pressure mercury lamp used with Fizeau fringes. Thus again on this count fringes of equal chromatic order can be obtained sharper than Fizeau fringes. A striking example of the definition obtainable by attention to the details discussed is shown in figure 4 of the plate. These are fringes of equal chromatic order (positive) for a calcite crystal matched against a specially prepared flat. The wave-lengths are indicated in ångströms. The value of t is 0·0015 mm. The silvering is very thick, R exceeding 0·90 and probably approaching 0·94 judging from the fringe-width. Exposure was a matter of some 15 minutes despite the thickness of the silver. The fringes have a striking narrowness. In the picture can be seen a small cleaved-out strip only 60 A. deep, yet clearly many times the amount that could still be resolved (the lattice spacing in calcite is 6 A.: see Tolansky and Khamsavi (1946)).

REFERENCES

Benoit, Fabry and Perot, 1913. *Bur. Intern.*, Vol. II.
Fabry, 1922. *Revue d'Optique*, **1**, 445.
Rasmussen, 1945. *Kgl. Danske Vidensk. Sel. Mat. Fys.*, **23**, 4.
Strong, 1940. *Modern Physical Laboratory Practice* (**Blackie**).
Tolansky, 1944 *Phil. Mag.*, **35**, 120.
Tolansky, 1945 a. *Proc. Roy. Soc.*, A, **184**, 41.
Tolansky, 1945 b. *Proc. Roy. Soc.*, A, **184**, 51.
Tolansky, 1945 c. *Phil. Mag.*, **36**, 236.
Tolansky, 1945 d. *Phil. Mag.*, **36**, 225.
Tolansky and Khamsavi, 1946. *Nature, Lond.*, **157**, **661**.
Tolansky and Wilcock, 1946. *Nature, Lond.*, **157**, 583.

A TRANSFORMATION OF KNOWN ASTIGMATISM FORMULAE

By H. H. HOPKINS,
W. Watson and Sons, Barnet, Herts

MS. received 26 *March* 1946

ABSTRACT. The use of effective " centres " and " radii " of curvature permits the well-known formulae relating the tangential and sagittal centres of curvature of a wave-front element before and after refraction to be put into a form which yields formulae and computing schemes identical with those customarily used for computing the paraxial focus, magnification, and equivalent focal length. Simple expressions follow for the *T*- and *S*-magnifications and oblique equivalent focal lengths.

After being in use for a few months, the computing schemes have now been adopted as standard calculations.

In figure 1, ACE′E is the optical axis of a lens system one of whose refracting surfaces, AP, is shown. Its pole is at A, its centre at C, and the radius AC $= r$. PE is a ray (continued) which is incident on the refracting surface at P, making an angle I with CP the normal at P. (For convenience the refracting surface is assumed to be spherical. The treatment can nevertheless be formulated for any surface of revolution.) The principal centres of curvature of an element of wave-front proceeding along this incident ray necessarily lie on the ray itself, since this is normal to the element of wave-front. The principal curvatures of the element of wave-front, assuming a ray in the meridian (tangential) plane, are known to be in the tangential and sagittal sections. The tangential section is the plane of the diagram in figure 1; and the sagittal section is a plane containing the ray and perpendicular to the plane of the diagram. Let T, S be the principal centres of curvature of the incident wave-front element; and let T′, S′ (lying

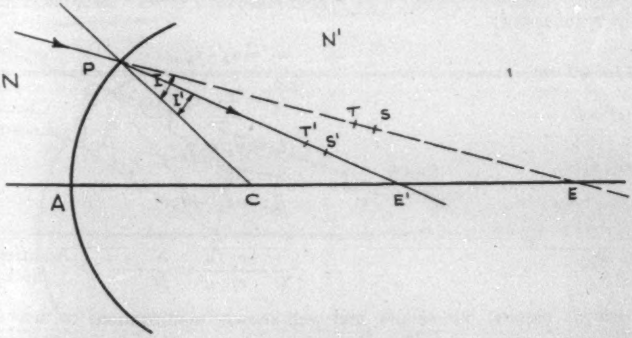


Figure 1.

on PE′, the refracted ray) be these centres for the refracted wave-front element which proceeds along PE′. This ray makes an angle I' with CP. Writing $\mathrm{PT}=t$, $\mathrm{PT}'=t'$, and $\mathrm{PS}=s$, $\mathrm{PS}'=s'$, the well-known formulae connecting these lengths are

$$\frac{N'\cos^2 I'}{t'}-\frac{N\cos^2 I}{t}=\frac{N'\cos I'-N\cos I}{r}, \qquad \ldots\ldots(1)$$

$$\frac{N'}{s'}-\frac{N}{s}=\frac{N'\cos I'-N\cos I}{r}. \qquad \ldots\ldots(2)$$

Each of these formulae is analogous to the paraxial formula

$$\frac{N'}{l'}-\frac{N}{l}=\frac{N'-N}{r} \qquad \ldots\ldots(3)$$

relating the centres of curvatures before and after refraction of an element of wave-front which proceeds along the optical axis of the system. Equation (3) can be looked upon as the degenerate form, common to both (1) and (2), when $I=I'=0$.

In lens computations, both for the purposes of calculating magnifications, equivalent focal lengths, and also for applying the sine condition and Herschel's condition, equation (3) is modified by the introduction of a paraxial incidence height at each surface y. (See Conrady (1929), whose notation is adopted except where stated to be otherwise.) The resulting formulae also afford useful check calculations.

In figure 2 is shown the diagram relating to an incident paraxial ray. The notation differs from Conrady's only in that $\mathrm{P\hat{C}A}=\alpha$. The computing scheme may be summarized:

$$\frac{l_k-r_k}{r_k}\,.\,u_k=i_k,$$

$$u_k+i_k=\alpha_k,$$

$$\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$$

$$\left.\begin{array}{l} y_k=r_k\,.\,\alpha_k, \\ y_k=y_{k-1}-d_{k-1}\,.\,u_k, \end{array}\right\}\ \text{Check values of } y_k$$

$$\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$$

$$i_k'=i_k\frac{N_k}{N_k'},$$

$$u_k'=\alpha_k-i_k',$$

$$\left.\begin{array}{l} l_k'=\dfrac{i_k'}{u_k'}\,.\,r_k+r_k, \\ l_k'=y_k/u_k', \end{array}\right\}\ \text{Check values of } l_k'$$

$$\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$$

$$l'_{k+1}=l_k'-d_k,$$

$$\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$$

$$\left.\frac{(l-r)\,.\,u}{(l'-r)\,.\,u'}=\frac{N'}{N}.\ \right\}\ \text{Additional check}$$

The scheme requires but simple, and well-known, modifications to meet the special cases $l=\infty$, $r=\infty$. The values of u_1, u_k' are used in finding the image size after the kth refraction, and in computing the coma as given by the sine-condition. They are also required in the computation of the change of spherical aberration consequent upon a change of object distance (see Hopkins, 1946).

Similar considerations can be applied to imagery along any ray, generally in practice a principal ray, which can then be looked upon as analogous to the axis. In what follows, it is shown that by introducing " effective " values of r, N, N', and a " paraxial " incidence height, there can be obtained formulae and computing schemes, of which that above is the degenerate case $I=I'=0$. In each of the calculations (for the T- and S-centres of curvature) values of u_1, u_k' are given which lead to " oblique " magnifications M_T, M_S and to " oblique " equivalent focal lengths F_T, F_S. The paraxial magnification and equivalent focal length again appear as the degenerate cases $I = I' = 0$.

In figure 3 is shown an incident ray PT together with a neighbouring ray P_TT (in the tangential plane) which intersect at T, the tangential centre of curvature of the wave-front element proceeding along PT. P_TT will be referred to as a T-ray and the point T as the T-focus; these are analogous to a paraxial ray and the paraxial focus. Draw a perpendicular from P_T on PT of length y; and draw CC_T perpendicular to PT to meet this ray in C_T. Then C_T will be looked upon as the effective " centre of curvature " of the refracting surface before refraction. The effective " radius of curvature " is $PC_T = r \cos I$. After refraction, C_T', lying on the refracted ray, and $PC_T' = r \cos I'$ are the effective " centre " and " radius " of the surface. Join now P_TC_T: then the triangle P_TC_TT corresponds

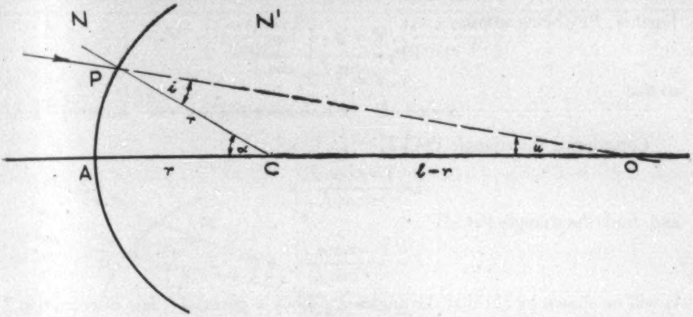


Figure 2.

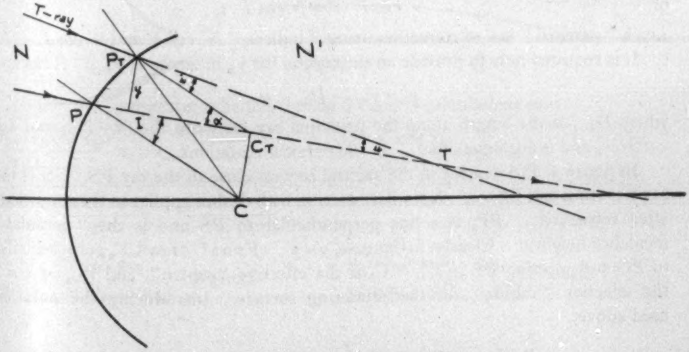


Figure 3.

with the triangle PCO of figure 2 and a corresponding notation has been introduced. It will be seen that

$$u=\frac{y}{t},\qquad \alpha=\frac{y}{r\cos I},\qquad \alpha=u+i;$$

$$u'=\frac{y'}{t'}\qquad \alpha'=\frac{y'}{r\cos I'},\qquad \alpha'=u'+i'.$$

Further, PP_T being small,

$$PP_T=\frac{y}{\cos I}=\frac{y'}{\cos I'},$$

so that

$$\alpha=\alpha',\quad \text{or}\quad u+i=\alpha=u'+i'.$$

Considering the triangle P_TC_TT,

$$i=\frac{t-r\cos I}{r\cos I}\,.\,u$$

and, from the triangle $P_TC_T'T'$,

$$i'=\frac{t'-r\cos I'}{r\cos I'}\,.\,u'.$$

It will be shown by (1) that the angles i, i' obey a paraxial " law of refraction " providing $N\cos I$, $N'\cos I'$ be written in place of N, N'.

Rearranging the terms in (1) gives, after multiplication of the left- and right-hand sides by $\dfrac{y}{\cos I}$, $\dfrac{y'}{\cos I'}$ respectively, gives

$$N'\cos I'\left\{\frac{t'-r\cos I'}{r\cos I'}\right\}.\,u'=N\cos I\left\{\frac{t-r\cos I}{r\cos I}\right\}.\,u$$

or

$$N'\cos I'\,.\,i'=N\cos I\,.\,i,$$

i.e., a " paraxial " law of refraction using " indices " $N'\cos I'$ and $N\cos I$.

It is required now to provide an expression for y_k in terms of y'_{k-1}. This is

$$y_k=y'_{k-1}-D_{k-1}\,.\,u_k,$$

where D_{k-1} is the length along the principal ray batween the $(k-1)$th and kth surfaces, and is analogous to d_{k-1} in the paraxial equations.

In figure 4, P_sS is a ray in the sagittal section close to the ray PS. S is the sagittal focus and $PS=s$. A similar notation with dashes applies to the conditions after refraction. PP_s is a line perpendicular to PS and is the " paraxial " incidence height y. Clearly, in this case, $y=y'$. From C draw CC_s perpendicular to PC and meeting PS in C_s. C_s is the effective "centre", and $PC_s=r/\cos I$ the effective " radius ", of the refracting surface. Introducing the notation used above,

$$u=\frac{y}{s},\qquad \alpha=\frac{y}{r/\cos I},\qquad \alpha=u+i;$$

$$u'=\frac{y}{s'},\qquad \alpha'=\frac{y}{r/\cos I'},\qquad \alpha'=u'+i'.$$

Considering the triangle P_sC_sS,

$$i=\frac{s-r/\cos I}{r/\cos I}\,.\,u$$

and, for the triangle $P_s'C_s'S'$,

$$i' = \frac{s' - r/\cos I'}{r/\cos I'} \, . \, u'.$$

It will be shown by (2) that i, i' obey a paraxial law of refraction, in which N, N' are the appropriate indices.

Rearranging the terms in (2) and multiplying by y gives

$$N' \, . \left\{ \frac{s' - r/\cos I'}{r/\cos I'} \right\} . \, u' = N \, . \left\{ \frac{s - r/\cos I}{r/\cos I'} \right\} . \, u$$

or

$$N' \, . \, i' = N \, . \, i,$$

i.e. a " paraxial " law using indices N, N'. As before,

$$y_k = y_{k-1} - D_{k-1} \, . \, u_k.$$

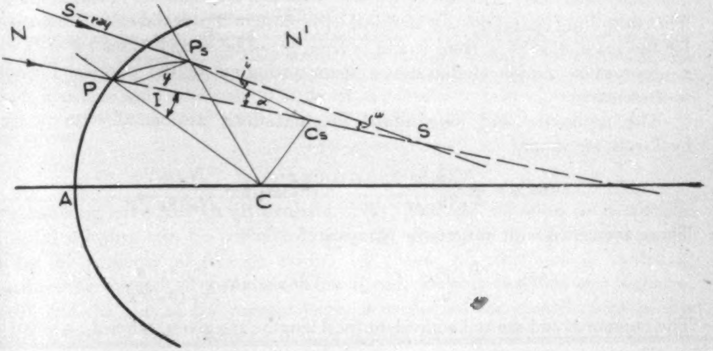


Figure 4.

Thus the computing schemes for the T- and S-calculations are:

$\dfrac{t_k - r_k \cos I_k}{r_k \cos I_k} \, . \, u_k = i_k,$	$\dfrac{s_k - r_k/\cos I_k}{r_k/\cos I_k} \, . \, u_k = i_k,$
$u_k + i_k = \alpha_k,$	$u_k + i_k = \alpha_k,$
. . . .	. . . .
$\left. \begin{array}{l} y_k = r_k \cos I_k \, . \, \alpha_k, \\ y_k = y'_{k-1} - D_{k-1} \, . \, u_k, \end{array} \right\}$ Check	$\left. \begin{array}{l} y_k = r_k/\cos I_k \, . \, \alpha_k, \\ y_k = y_{k-1} - D_{k-1} \, . \, u_k, \end{array} \right\}$ Check
. . . .	. . . .
	$\alpha_k' = \alpha_k \cdot \dfrac{\cos I_k'}{\cos I_k},$
$i_k' = i_k \cdot \dfrac{N_k \cos I_k}{N_k' \cos I_k'},$	$i_k' = i_k \cdot \dfrac{N_k}{N_k'},$
$u_k' = \alpha_k - i_k',$	$u_k' = \alpha_k' - i_k',$
. . . .	. . . .

$$\left.\begin{aligned} t_k' &= \frac{i_k'}{u_k'} \,.\, r_k \cos I_k' + r_k \cos I'_k, \\ y_k' &= y_k \,.\, \frac{\cos I_k'}{\cos I_k}, \\ t_k' &= y_k'/u_k', \end{aligned}\right\}\text{Check} \qquad \left.\begin{aligned} s_k' &= \frac{i_k'}{u_k'} \,.\, r_k/\cos I_k' + r_k/\cos I_k', \\ \\ s_k' &= y_k/u_k', \end{aligned}\right\}\text{Check}$$

$$\begin{matrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \qquad & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}$$

$$t_{k+1} = t_k' - D_k, \qquad\qquad s_{k+1} = s_k' - D_k,$$

$$\begin{matrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \qquad & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}$$

Additional checks:—

$$\frac{t - r\cos I}{t' - r\cos I'} = \frac{N'}{N} \qquad\qquad \frac{s - r/\cos I}{s' - r/\cos I'} = \frac{N'\cos I'}{N\cos I}.$$

The computing operations are identical with those of a paraxial calculation except for the calculation of y' from y, and α' from α. The special cases $t = \infty$, $s = \infty$, $r = \infty$ require simple modifications, again analogous to the paraxial formulae in these cases.

The transverse and longitudinal magnifications associated with imagery by T-rays are simply

$$(M)_T = \frac{N_1 (u_1)_T}{N_k' (u_k')_T}, \qquad (M_{\text{long}})_T = \frac{N_1 (u_1)^2{}_T}{N_k' (u_k')^2{}_T}.$$

Those associated with imagery by S-rays are

$$(M)_S = \frac{N_1 (u_1)_S}{N_k' (u_k')_S}, \qquad (M_{\text{long}})_S = \frac{N_1 (u_1)^2{}_S}{N_k' (u_k')^2{}_S}.$$

The tangential and sagittal equivalent focal lengths are given (when t_1, $s_1 = \infty$) by

$$F_T = \frac{y_1}{(u_k')_T}, \qquad F_S = \frac{y_1}{(u_k')_S}.$$

In these magnifications and equivalent focal length formulae the object and image elements are assumed perpendicular to the incident and emergent (principal) rays.

The forms of sine condition and Herschel's condition applicable to oblique imagery can be readily formulated by a method previously given for axial imagery (see Hopkins, 1946 a, b).

REFERENCES

CONRADY, A. E., 1929. *Applied Optics and Optical Design* (London: Oxford University Press), Chapter I.
HOPKINS, H. H., 1946 a. *Proc. Phys. Soc.*, 58, 92.
HOPKINS, H. H., 1946 b. *Proc. Phys. Soc.*, 58, 100.

THE MECHANICAL PROPERTIES OF METALLIC SOLID SOLUTIONS

By F. R. N. NABARRO,
Royal Society Warren Research Fellow

Communicated by N. F. Mott, F.R.S. MS. received 28 *May* 1946

ABSTRACT. The theoretical relation between the lattice strains produced by precipitation in a metal and the corresponding increase in hardness is extended to the case of lattice strains in metallic solid solutions. The elastic limit of a single crystal of a solid solution is calculated on the assumption that the crystal will slip when the applied external stress is equal to the mean value of the internal stress. This mean is taken over the length of a dislocation, assumed to be 1000 atoms. The estimate agrees in order of magnitude with the experimental observations. Similar considerations are applied to the hardness of polycrystalline solid solutions. The theory is extended to cases in which the increase of hardness produced by alloying is not large in comparison with the hardness of the pure solvent.

§1. INTRODUCTION

It has long been known (Rosenhain, 1921) that the hardness of a metallic solid solution can be correlated with the distortion of the lattice caused by the presence of foreign atoms. A theory of precipitation hardening based on the concept of dislocations has already been given (Mott and Nabarro, 1940), and the aim of the present paper is to discuss the modifications in this theory which become necessary if it is applied to solid solutions in which the nuclei of strain are single atoms rather than aggregates of many atoms. These modifications are of two kinds. Firstly, it is no longer justifiable to estimate the strain in the matrix from purely elastic considerations of the atomic volumes of solvent and solute, and account must be taken of the binding forces between atoms of the solvent and of the solute. Secondly, even in solutions as dilute as 1%, the internal strains caused by the dissolved atoms change their directions over distances of a few lattice spacings, and it is to be expected that the opposed stresses acting on closely adjoining parts of a dislocation will largely cancel one another and so reduce the applied stress necessary to move the dislocation. It is apparently this effect which accounts for the great increase in mechanical strength produced by precipitation hardening.

§2. THE HARDNESS OF AN ALLOY CONTAINING A PRECIPITATED PHASE

According to the theory advanced by Mott and Nabarro, the elastic limit S of an alloy hardened by the precipitation of a second phase in roughly spherical particles should be given in order of magnitude by

$$S = E\epsilon f, \qquad \ldots\ldots(1)$$

where E is Young's modulus for the matrix (which to the order of accuracy required may be taken as the modulus for the pure solvent), f is the concentration of the precipitated phase, and ϵ measures the misfit of the precipitate in the lattice of the matrix when the stresses in the matrix and in the precipitate are in equilibrium.

If this formula is applied to solid solutions, S and E retain their meanings unchanged, and f is the atomic concentration of the solute. The effective value of ϵ may be obtained directly from the difference of the measured x-ray lattice parameters of the solid solution and of the solvent. If the lattice parameter of the solvent is a and that of the solution is $a+\Delta a$, the fractional change in atomic volume due to the presence of the solute is $(1+\Delta a/a)^3-1$. This fractional change is also given by $(1+\epsilon f)^3-1$, and the relation between ϵ and Δa is thus

$$\epsilon f=\Delta a/a. \qquad \ldots\ldots(2)$$

§ 3. INFLUENCE OF THE SMALL SCALE OF THE STRAINS

In its application to precipitation hardening, the theory of Mott and Nabarro (1940) assumed that the regions in which the strain retained a roughly constant value were so large that an external stress could not cause a dislocation to move until it reversed the strain in each region through which the dislocation passed. Dislocations, pictured as straight lines since the early theories (Taylor, 1934; Orowan, 1934; Polanyi, 1934), must in fact be to some extent flexible, as has been emphasized by Burgers (1940). In accordance with the ideas of Mott (1946), the energy of a dislocation is an increasing function of its length, and this rate of change of energy with length leads to properties of the dislocation analogous to those of a soap film. A dislocation can assume a wavy form in which its radius of curvature is everywhere large compared with atomic dimensions under the influence of stresses small compared with the elastic constants of the material. In a precipitation-hardened alloy in which the stresses retain the same sign over distances of hundreds of atomic spacings, the equilibrium position of a dislocation is one in which each part of the dislocation lies in a region of the crystal where its elastic energy is low. To move such a dislocation involves lifting all parts simultaneously from potential troughs, and so requires an external stress $E\epsilon f$. Under small-scale stresses, such as are produced by solution hardening, the dislocation cannot follow the contours of minimum elastic energy, for this would require it to adopt a form in which its radius of curvature was of atomic dimensions, and so would require stresses of the order of the elastic constants. If the axis of a dislocation were a rigid straight line, the stress opposing its motion would vary in sign from region to region along its length and the resultant force would be small. For example, in a solid solution of atomic concentration f, the amplitude of the stress is of the order $E\epsilon f$, and this stress changes sign at the boundaries of regions having linear dimensions of the order of $f^{-\frac{1}{3}}$ atomic spacings. The expected mean value of such an oscillating function depends on the size of the region over which the mean is taken. The length of a dislocation is not known, but the fact that the yield stress of a single crystal is independent of the size of the specimen suggests that slip takes place independently in domains which are small compared with the dimensions of the specimen. The size of these domains is not known, but they may tentatively be identified with the mosaic elements of linear dimensions 10^{-4} or 10^{-5} cm. which are believed to form a real crystal. The stress required to move a straight dislocation 1000 atoms long is the mean of $1000f^{\frac{1}{3}}$ contributions of random sign and of amplitude $E\epsilon f$, and so is of order $E\epsilon f/\surd(1000f^{\frac{1}{3}})$. In a

5% solution, with $f=1/20$, this is about 1/20 of $E\epsilon f$. It is possible that the mean stresses may be even smaller, for the solution in its state of lowest energy will tend to have foreign atoms as far from one another as possible, and so arranged at roughly equal distances apart. In the extreme case of an ordered solid solution, the components of stress along successive elements of the dislocation of individual length $f^{-\frac{1}{3}}$ lattice spacings cancel, and the resultant mean stress on a straight dislocation is of order only $E\epsilon f/1000f^{\frac{1}{3}}$, or 1/400 of $E\epsilon f$. In many order-disorder transformations the disordered phase is cubic and the ordered phase is of lower symmetry. Within a single crystal, ordering starts from many nuclei, and the crystal in the ordered state is broken into domains differing in the orientation of their lattices. The resulting large-scale internal strains produce hardening effects comparable to those produced by precipitation (Barrett, 1943). However, in the case of β brass (Smith, 1942), in which the ordered lattice is cubic, annealing at room temperature produces a softening which has been attributed to the growth of the domains of coherent order.

According to these arguments, the elastic limit of a dilute (and therefore disordered) solid solution should be given by an expression of the form

$$S=\alpha E\epsilon f, \qquad \ldots\ldots(3)$$

where S, E, ϵ and f have their previous meanings, and α is a pure number equal to about 0·05, and almost independent of the concentration.

§4. COMPARISON OF THE THEORETICAL FORMULA WITH OBSERVATIONS

(i) *Au in Ag, Ag in Au.*

For this system the resolved shear stress across the glide plane in the glide direction is given by Sachs and Weerts (1930) in the form

$$\tau=0{\cdot}060k+0{\cdot}091(1-k)+1{\cdot}708k(1-k),$$

where τ is the stress in kg./mm^2, and k is the atomic concentration of Ag.

They also give the lattice parameter in the form

$$a=4{\cdot}07765k+4{\cdot}0700(1-k)-0{\cdot}240k(1-k).$$

Taking $E=8\times10^{11}$ dynes/cm^2 for Au, $E=7{\cdot}7\times10^{11}$ for Ag, and substituting the values of $d\tau/dk$ and da/dk given by the above expressions when $k=0$ and when $k=1$ into equations (2) and (3) leads to $\alpha=0{\cdot}051$ for Ag dissolved in Au, $\alpha=0{\cdot}028$ for Au dissolved in Ag.

(ii) *Zn in Cu*

Here von Göler and Sachs (1929) give an elastic limit which initially rises linearly with the concentration of Zn, the slope of the stress-concentration curve being 0·1 kg./mm^2 for 1% addition of Zn. The lattice parameter rises linearly with concentration from $a=3{\cdot}6088$ A. for pure Cu to $a=3{\cdot}686$ A. for 67·5% Cu. These values give $\alpha=0{\cdot}015$.

(iii) *Zn in Al*

For this system Elam (1925) says that single crystals of Al + 18·6% Zn show a well-marked yield at 15 tons/sq. in. (which would correspond to a resolved stress of about 7½ tons/sq. in.). Accepting Elam's figures for the lattice parameters (Al 4·06 A., alloy 4·18 A.), and taking E for Al as 7×10^{11} dynes/cm^2 would lead to $\alpha = 0{\cdot}056$. However, a large expansion of the lattice on adding Zn to Al is unexpected, and is not supported by the results of other workers. If Elam's figures for the lattice parameter are replaced by those of Owen and Iball (1934) (Al: 4·041 A., Al + 19% Zn: 4·0345 A.), the value of α becomes 1·05. Such a large value should be characteristic of precipitation hardening rather than solution hardening. The solubility of Zn in Al is not known beyond dispute, but Hansen (1936) considers the most reliable results to be those of Schmid and Wassermann, leading to a solubility of less than 2% at room temperatures. This gives considerable support to the suggestion that Elam's crystals were age-hardened.

(iv) *Cd in Zn*

In this system again the hardening found by Rosbaud and Schmid (1925) is remarkably great. Their values for the elastic limit may be combined with the value $E = 8 \times 10^{11}$ dynes/cm^2 for Zn. The value of ϵ may be estimated from the atomic volumes as $(13{\cdot}01 - 9{\cdot}16)/3 \times 9{\cdot}16$, and this leads to $\alpha = 0{\cdot}22$. This high value might be attributed to some peculiarity of the hexagonal lattice, but such an explanation seems unlikely, for in a single crystal specimen both hexagonal and face-centred cubic metals are known to slip on just one family of close-packed planes. A more likely explanation is that the specimens were precipitation hardened, since Hansen accepts the estimate of Boas that the solubility of Cd in Zn is less that 0·4% at 150° C. and unmeasureably small at 100° C. or below. Rosbaud and Schmid assume that the Cd is in solution because no eutectic is visible in the microstructure. On the other hand, the microphotographs of single crystals reproduced show large inclusions of some other phase or orientation. These may be twins (which would harden a single crystal of a hexagonal metal considerably), but are more likely to be laminar precipitates of another phase formed by diffusion.

(v) *Ni in Cu, Cu in Ni*

This system was investigated by Osswald (1933). The curve relating elastic limit and concentration is distorted from the simple parabolic form which represents the results for the system Au–Ag. Osswald finds a negligibly low elastic limit for copper, rising to a maximum resolved shear stress of 3·0 kg./mm^2 at a concentration of about 60 atomic % Ni, and falling to 0·6 kg./mm^2 for pure Ni. For dilute solutions, the slopes of the curves may be estimated as 0·2 kg./mm^2 for 1% Ni added to Cu, and 0·17 kg./mm$_2$ for 1% Cu added to Ni. The Young's moduli are 11×10^{11} dynes/cm^2 for Cu and 21×10^{11} dynes/cm$_2$ for Ni. The lattice parameters are $a = 3{\cdot}608$ A. for Cu, $a = 3{\cdot}5175$ A. for Ni (Burgers and Basart, 1930; Owen and Pickup, 1934). Taking the mean of the results of the authors quoted, Δa for 1% addition is $-0{\cdot}0013$ A. for Ni added to Cu, 0·00078 A. for Cu added to Ni. These values substituted in equations (2) and (3) give $\alpha = 0{\cdot}049$ for Ni dissolved in Cu, and $\alpha = 0{\cdot}036$ for Cu dissolved in Ni.

(vi) *Cu in Au, Au in Cu*

The elastic limits of polycrystalline alloys of copper and gold were determined by Broniewski and Wesolowski (1934). 1 atomic % Cu added to Au increases the elastic limit by 0·5 kg./mm^2, and 1 atomic % Au added to Cu increases the elastic limit by 0·7 kg./mm^2 The corresponding values of α are 0·073 and 0·066.

§5. APPLICATION TO HARDNESS MEASUREMENTS

A number of workers have measured the hardness of solid solutions (Goebel, 1922; Norbury, 1924; Frye and Hume-Rothery, 1942; Brick, Martin and Angier, 1943), and an interpretation of their results on the lines of the present theory would be of interest. However, in the Brinell test there is very heavy cold working of the metal close to the indentation: Norbury (1924) and Frye and Hume-Rothery (1942), by taking as their measure of hardness the " ultimate Meyer hardness ", measure a quantity which lies even farther from the initial elastic limit and closer to the limiting strength after heavy cold work than the Brinell number determined from a shallow impression. Attempts to extrapolate Frye and Hume-Rothery's results by Meyer's formula to give the hardness corresponding to an impression of diameter 1/10 of that of the ball lead to very irregular figures, as was to be expected from the work of Norbury (1923) on annealed copper and copper alloys. It appears that the experimental results measure the limiting strength after heavy cold work rather than the elastic limit. In general, dilute solid solutions work-harden under small strains less rapidly than does the pure solvent. Elam (1927) finds that in single crystals of brass this is also true for large concentrations and large strains, but Brick, Martin and Angier (1943) find that rolling reductions of 20% or more produce increases in the hardness of polycrystalline solid solutions in copper which are linearly proportional to the excess of the hardness of the annealed alloy above that of pure copper.

The reason for the abnormally small work-hardening of solid solutions under small strains is not known. It is presumably connected with the fact observed by Masima and Sachs (1928) that, whereas in single crystals of pure metals the active glide plane gives way to a new plane as soon as the resolved stress on the new plane exceeds that on the active plane, in solid solutions the original plane remains active until the stresses on the new plane considerably exceed those on the original plane. Possibly the slip process itself produces a type of ordering of the solid solution which is conducive to slip on the active plane. An alternative possibility is that the initial irregularity of the lattice of the solid solution causes small areas of lattice planes to buckle under small stresses, and so prevents the catastrophic rotations of large blocks of the lattice which lead to Laue asterisms in the early stages of slip in pure metals. It would be of interest to know if the early stages of slip in solid solutions, which are characterized by abnormally small hardenings, do in fact show abnormally little lattice distortion in Laue photographs.

Owing to this complicating factor, the analysis of hardness tests is not straightforward. It will be assumed that the Brinell or ultimate Meyer hardness exceeds the elastic limit of a single crystal by a factor β, so that the hardness is given by

$$H = \alpha\beta E\epsilon f. \qquad \ldots\ldots(4)$$

The numerical value of β is not known with any accuracy, nor (except in special cases) is it certain that the value of β remains roughly constant for a series of solutions in the same solvent. A rough estimate of the magnitude of β may be based on the consideration that in general the Brinell hardness of a pure annealed metal is of order 40 kg./mm.^{2} (e.g. Fe 67, Ag 30, Cu 45, Al 16), while the elastic limit for a single crystal is of order 0·2 kg./mm^2 (e.g. Ag 0·06, Cu 0·1–0·4, Al 0·5), giving β about 200. Solid solutions showing little work-hardening would yield smaller values of β, and this estimate may be combined with one derived from a comparison of the measurements of Rosbaud and Schmid on the yield points of alloys of zinc and cadmium with the hardness measurements of Ludwik, which they quote. The two curves are not very similar in form, but the mean value of β is about 10. A figure of 80 will be taken as an estimate of the order of magnitude of β.

§ 6. SUMMARY OF THE EXPERIMENTAL OBSERVATIONS ON HARDNESS

Norbury (1924) examined the hardness of solid solutions of Ni, Zn, Al, Mn, Ag and Sn in copper, and similarly analysed the results of Goebel (1922) on solid solutions of Bi, Sn, Hg, and Cd in lead. In both cases the hardening effect is very roughly a linear function of the difference in size of the atoms of the solute and the solvent. In each series of experiments there was one anomaly, the solid solutions of Si in Cu and Na in Pb both being abnormally hard. Norbury attributes this to the formation of intermetallic compounds, so that the dissolved atom no longer has twelve symmetrically-placed nearest neighbours, but is closely bound to a smaller number of solvent atoms. This introduces stresses into the lattice which are more complicated than those which occur in other cases. Similar results have been obtained by Frye and Hume-Rothery (1942) on the hardness of solid solutions of Cd, In, Sn, Sb Mg, Al and Zn in silver. The solution of Al and Zn in silver decreases the lattice constant, whereas the other solutes increase the lattice constant. In all cases the hardness of solid solutions of various concentrations of these seven elements in silver is roughly a linear function of the change in lattice spacing (without regard to sign). Solutes which decrease the lattice spacing appear in general to produce a greater increase in hardness for a given change of lattice spacing than those which increase the lattice spacing. Frye and Hume-Rothery show that if the observations are grouped into series in which the dissolved atoms belong to the same row of the periodic table, and are present in the same concentration (2·4 atomic per cent or 5·0 atomic per cent), then the hardness in each series is very closely given by a linear function of the square of the change in lattice constant. The disagreement of the square law with the linear law originally suggested is discussed below. Finally, results by Brick, Martin and Angier (1943) for Si, Ni, Zn, Al, Mn, As, Sb and Sn dissolved in copper largely confirm the work of Norbury. Cobalt has an atomic volume slightly less than that of copper, and its solid solutions are very hard. The change in lattice parameter caused by adding Mg is not known, but the point for Mg falls on the curve relating hardening effect and difference of atomic volume.

A close connection between hardness and internal stress has recently been demonstrated by Frye, Caum and Treco (1944).

§ 7. EFFECT OF RESIDUAL STRESSES DUE TO OTHER CAUSES

The theory originally proposed led to a hardness of the alloy directly proportional to the deformation of the lattice, whereas the measurements of Frye and Hume-Rothery show a linear relation between hardness and the square of the deformation. These are in fact both limiting cases of the same general formula, and are obtained when the increase in hardness caused by the presence of the dissolved atoms is very large or very small in comparison with the hardness of the pure solvent.

Whether the dislocation is imagined as rigid or as flexible, the stress required to cause slip in a metal by moving an appreciable number of dislocations in it is determined less by the extreme values of the internal stress than by the average fluctuations of stress in the material. These are conveniently measured by the mean square stress, since the mean stress over the whole body must be zero.

Consider a pure metal containing random internal stresses (due to misfits at grain boundaries or imperfect annealing) of mean square value S_0^2. The hardness of this metal is given by

$$H_0 = \alpha\beta S_0, \qquad \ldots\ldots(5)$$

where α is a pure number, the value of which, according to the previous arguments, must be about 1 if dislocations are flexible and about 1/20 or less if dislocations are rigid, and β is the ratio of the hardness to the elastic limit, and is of order of magnitude 50 to 100. If foreign atoms are now dissolved in the metal they give rise to a further random distribution of stress of mean square value $(E\epsilon f)^2$. The hardness of the solid solution is now given by

$$H = \alpha\beta\sqrt{(S_0^2 + E^2\epsilon^2 f^2)}. \qquad \ldots\ldots(6)$$

Eliminating S_0 between (5) and (6) gives

$$H^2 = H_0^2 + (\alpha\beta E\epsilon F)^2. \qquad \ldots\ldots(7)$$

The original theory, which neglected H_0, made H a linear function of $E\epsilon f$, and Goebel's results for solutions in lead, where the initial hardness is small, show a roughly linear dependence of hardness both on concentration (except where the solute is Hg) and on difference of atomic volume (except in the case of Na, which has already been attributed to the formation of a chemical bond).* Norbury's solid solutions in copper and Frye and Hume-Rothery's solutions in silver showed only small increases in hardness above that of the pure solvent. For such cases equation (6) may be written approximately as

$$H - H_0 = (\alpha\beta E\epsilon f)^2/2H_0. \qquad \ldots\ldots(8)$$

Norbury's results show an increase in hardness which for dilute solutions of a given element varies more rapidly than a linear function of the concentration in the cases of solutions of Si, Al, Ni and Zn. The effect is not shown in solutions of Sn, Ag, and Mn, but in the cases of Sn and Mn he states that " there are reasons for thinking that . . . . the results really lie on a curve bending upwards initially ", and in the case of Ag " it is possible that . . . . the results lie on a curve bending upwards and not on a straight line ". The results of Brick, Martin and Angier on 1% solutions in copper are well represented (except for Si) as a function of ϵ by a formula of type (7), and finally the quadratic law of Frye and Hume-Rothery is in agreement with the form of equation (8).

However, the dependence of hardness on concentration in the results of Brick, Martin and Angier for solutions in copper is linear even for small concentrations which increase the hardness by an amount less than the initial hardness of pure copper, and it seems in general that the hardening effects of solution and of other causes combine sometimes linearly and sometimes as independent random factors. It is to be expected that the small-scale strains due to solution will combine in a random manner with other small-scale strains, but will add linearly to the hardness produced by large-scale strains.

---

* Kurnakow *et al.* (1940) have confirmed that small additions of Na harden lead appreciably, while altering the lattice parameter very little.

§8. ANALYSIS OF EXPERIMENTAL HARDNESS MEASUREMENTS

As a check on the order of magnitude of the effect, the quantity $\alpha\beta$ of equation (8) may be estimated from the results of Frye and Hume-Rothery. They give $H_0=43$ kg./mm.$^2=4{\cdot}2\times10^9$ dynes/cm.^{2} and E for silver is $7{\cdot}7\times10^{11}$ dynes/cm.^{2} The lattice spacing a in silver is $4{\cdot}1$ A., and so $\epsilon f=\Delta a/4{\cdot}1$. For $f=5\%$. Frye and Hume-Rothery give two lines, one of slope $\Delta H/(\Delta a)^2=2{\cdot}4\times10^5$ kg./mm.^{2}/A.^{2}, the other of slope $0{\cdot}45\times10^5$ kg./mm.^{2}/A^2. (The factor 10^5 appears in their paper as 10^7, but this is an error in computation.) These correspond respectively to $\Delta H/\epsilon^2 f^2 = 4\times10^{14}$ dynes/cm.^{2} and $0{\cdot}74\times10^{14}$ dynes/cm.^{2} So the value of $\alpha\beta=(2H_0\Delta H/\epsilon^2 f^2)^{\frac{1}{2}}/E$ is $2{\cdot}4$ or $1{\cdot}0$.

It is also possible to estimate $\alpha\beta$ from Goebel's results on solutions in lead, which give hardening as a linear function of the distortion of the lattice. Norbury's plot of Goebel's results is expressed in arbitrary units, which appear to be hardening h in tenths of a kilogram per square millimetre and difference of size of solvent and solute atoms e in cubic centimetres per gram atom. This gives $H=0{\cdot}981\times10^7 h$ dynes/cm.^{2}, $\epsilon=e/3\times18{\cdot}27$. For $f=0{\cdot}01$, the points plotted by Norbury show $h/e=2{\cdot}7$, corresponding to $H/\epsilon=1{\cdot}45\times10^9$ dynes/cm.^{2} Taking E for lead as $1{\cdot}6\times10^{11}$ dynes/cm.^{2} gives $\alpha\beta=H/E\epsilon f=0{\cdot}9$. This value of $\alpha\beta$ is consistent with that deduced from the results of Frye and Hume-Rothery. Combining these results with the rough estimate $\beta=80$ gives values of α of the order of $0{\cdot}02$, in agreement with the estimates from the yield points of single crystals.

ACKNOWLEDGMENT

The writer is indebted to Professor N.-F. Mott, F.R.S. for many stimulating discussions on this problem.

REFERENCES

BARRETT, C. S., 1943. *Structure of Metals* (McGraw-Hill).

BRICK, R. M., MARTIN, D. L. and ANGIER, R. P., 1943. *Trans. Amer. Soc. Met.*, **31**, 675.

BRONIEWSKI, W. and WESOLOWSKI, K., 1934. *C.R. Acad. Sci., Paris*, **198**, 569.

BURGERS, J. M., 1940. *Proc. Phys. Soc.*, **52**, 23.

BURGERS, W. G. and BASART, J. C. M., 1930. *Z. Kristallogr.*, **75**, 155.

ELAM, C. F., 1925. *Proc. Roy. Soc.*, A, **109**, 143; 1927. *Ibid.*, **115**, 148.

FRYE, J. H., CAUM, J. W. and TRECO, R. M., 1943. *Trans. Amer. Inst. Min. Met. Engrs.*, **152**, 83.

FRYE, J. H. and HUME-ROTHERY, W., 1942. *Proc. Roy. Soc.*, A, **181**, 1.

GOEBEL, J., 1922. *Z. Metallk.*, **14**, 357, 388, 425, 449.

GÖLER, FRHR. v. and SACHS, G., 1929. *Z. Phys.*, **55**, 581.

HANSEN, M., 1936. *Aufbau der Zweistofflegierungen* (Berlin : Springer).

KURNAKOW, N. S. *et al.*, 1940. *Ann. Secteur Anal. Phys.-Chim. U.S.S.R.*, **13**, 233

MASIMA, M. and SACHS G., 1928. *Z. Phys.*, **50**, 161.

MOTT, N. F., 1946. *J. Inst. Metals*, **72**, 367 (May Lecture).

MOTT, N. F. and NABARRO, F. R. N., 1940. *Proc. Phys. Soc.*, **52**, 8.

NORBURY, A. L., 1923. *J. Inst. Metals*, **29**, 407, 423.

NORBURY, A. L., 1924. *Trans. Faraday Soc.*, **19**, 586.

OROWAN, E., 1934. *Z. Phys.*, **89**, 605, 614, 634.

OSSWALD, E., 1933. *Z. Phys.*, **83**, 55.

OWEN, E. A. and IBALL, J., 1934. *Phil. Mag.*, **17**, 433.

OWEN, E. A. and PICKUP, L., 1934. *Z. Kristallogr.*, **88**, 116.

POLANYI, M., 1934. *Z. Phys.*, **89**, 660.

ROSBAUD, P. and SCHMID, E., 1925. *Z. Phys.*, **32**, 197.

ROSENHAIN, W., 1921. *Proc. Roy. Soc.*, A, **99**, 196.

SACHS, G. and WEERTS, J., 1930. *Z. Phys.*, **62**, 473.

SMITH, C. S., 1942. *Metals Technol.*, T.P. 1517.

TAYLOR, G. I., 1934. *Proc. Roy. Soc.*, A, **145**, 362.

ON METHODS FOR THE CONSTRUCTION OF NETWORKS DUAL TO NON-PLANAR NETWORKS

By A. BLOCH

(Communication from the Staff of the Research Laboratories of
The General Electric Company Limited, Wembley, England)

MS. received 23 *August* 1945 ; *in revised form* 27 *June* 1946

ABSTRACT. Two networks are said to be dual to each other if the following correspondence exists between them :—

(*a*) Each element of one network has a counterpart in the other one.

(*b*) The currents through the elements in one network and the voltages across their counterparts in the other network are proportional to each other.

A network dual to a given one is usually constructed by connecting elements of the appropriate kind in a circuit the diagram of which forms the topological counterpart to the diagram of the original circuit. If this original circuit is a " non-planar " one (i.e. if it cannot be drawn without crossings between some of its branches), then there exists no topological dual, and for this reason it has been said that there exists no dual to a non-planar electrical network. However, if we are interested in an electrical network not because its circuit diagram has certain geometrical properties, but because it constitutes an assembly of a number of passive and active elements which perform in a certain manner (on account of their impedance properties and of certain relations of mutual constraint imposed by Kirchhoff's mesh and junction relations), then there is really no restriction of this kind.

We need only convert the original network into another with the same elements and the same performance, but with a planar circuit diagram and with additional constraints implemented by ideal transformers. This equivalent network can be then converted into its dual counterpart, which is another planar network subject to the dual counterparts of the transformer constraints of the first case. In this way we arrive at an assembly which performs in the required dual manner.

Introducing thus the concept of physical duality as opposed to geometrical duality, the paper explains in an elementary manner a number of ways in which this conversion into equivalent planar networks can be achieved. These methods are :—

(1) A method originally published by Julia.

(2) The dual counterpart of this method.

(3) The *method of fictitious junction points* (an electrical analogue to F. Schur's method of fictitious junction points in pin-jointed frameworks).

(4) The *method of the transferred terminal*.

(5) The general *method of separation*.

Some of these methods require a smaller number of ideal transformers than Julia's original method.

The paper ends with a short discussion of the errors introduced when the dual network is set up on a network analyser and when the ideal transformers have to be replaced by real ones ; the numerous methods available for the realization of an equivalent network enable a choice to be made so as to minimize these errors, or possibly eliminate them altogether.

The existence of a dual to a non-planar network is of some interest in connection with electro-mechanical analogies as it removes a restriction to which the " direct " analogy would otherwise be subjected.

§1. INTRODUCTION: THE PRINCIPLE OF DUALITY

Two networks are said to be dual to each other if the following correspondence exists between them:—

(*a*) Each element of one network has a counterpart in the other.

(*b*) The currents through the elements of one network and the voltages across their counterparts in the other network are proportional to each other.

In other words, we have the following relations:—

$$i_\kappa' R_0^1 = e_\kappa'' , \qquad \ldots\ldots(1)$$

$$i_\kappa'' R_1^0 = e_\kappa' , \qquad \ldots\ldots(2)$$

where i_κ' and e_κ' are current and voltage at element κ' of the first network and i_κ'' and e_κ'' current and voltage at the corresponding element κ'' of the second network, and where R_0^1 and R_1^0 are two constants (each of the dimension of a resistance) which are characteristic of the particular correspondence under consideration.

It is obvious then that the behaviour of one network is a copy of that of the other network, but with the rôles of current and voltage interchanged.

We shall see in the following that the existence of such a relation can be predicted by very simple criteria, and for this reason it will obviously be useful, for it will enable us to predict the performance of one network without a preceding analysis provided we know the behaviour of its dual counterpart. This *Principle of Duality* has been known for some time. The phrase seems to have been introduced into electric network theory by H. Sire de Vilar (1901). The first textbook dealing with it was A. Russell's *Theory of Alternating Currents* (1904), which gives a large number of applications and a list of earlier references. There are, or rather were, however, certain difficulties in the application of the principle whenever one of the networks concerned was " non-planar."*

These difficulties give rise to an interesting question in connection with electro-mechanical analogies (Bloch, 1945). There are two types of such an analogy: one in which we represent a force by a voltage (the " direct " analogy), and one in which we represent a force by a current (the " inverse " analogy). If we represent a given mechanical system once by the first and once by the second type of analogy then we obtain two electrical networks which are—obviously—in dual relationship to each other. Now the circuit diagram of the inverse type of analogy can always be obtained as a simple translation of the mechanical layout if this is drawn according to certain conventions. Should this circuit diagram be of the non-planar type, and were we to take the statement literally that there exists no dual to such a network,† then we would have to acknowledge that the field of application of the direct analogy was restricted. We shall see that this is not the case, and that the apparent contradiction turns on the definition of the concept of duality. Before concentrating on the special problem of the present paper we shall therefore deal first more fully with the normal and general aspects of the duality principle.

---

* A non-planar network is one the circuit diagram of which cannot be drawn without crossings between some of its branches.

† Whitney (1931) shows that there exists no geometrical dual to a non-planar graph; Barnes and Gardener (1932) seem to take this as the basis of their statement that there exists no dual to a non-planar network.

It will be seen that we have to distinguish between duality in a topological sense and duality in a physical sense. By duality in a physical sense we have in mind a correspondence between two assemblies of elements such as outlined above, a correspondence which is concerned only with the behaviour of the elements specified. By duality in the topological sense we have in mind a relation between the structure of the two circuit diagrams such that each branch of one network has its counterpart in the other network and that all those branches of one network which radiate from a particular junction point have counterparts in the other network which form a closed mesh. The fact that the branches of the network are linked together in a certain way imposes on the currents therein and on the voltages across them certain conditions of constraint commonly known as Kirchhoff's equations. These conditions of constraint are obviously decisive for the behaviour of the network. Now we shall see that in order to bring about dual behaviour of two networks it is sufficient to establish a certain impedance relation between corresponding elements of the two networks, to provide external sources of dual character, and finally to provide conditions of constraint which are of dual character. If we insist in bringing about this dual type of constraint by an appropriate structure of the circuit diagram—and by this means only—then we shall see that the concepts of physical and of topological duality are inseparably linked. On the other hand, if we are only interested in the behaviour of the networks (i.e., assemblies of elements), such topological duality is immaterial; we may then introduce the required conditions of constraint by other means. These other means are ideal transformers, and we shall see by their use it will always be possible to specify one or more networks which are dual in the physical sense. Thus the theorem that there exists no topological dual to a non-planar network does not occupy the dominant position which one might at first assume. It is nevertheless still of interest, for obviously topological duality is the simplest way of bringing about dual conditions of constraint, and we shall always try first to achieve it. Knowledge of the theorem will, however, save us from useless attempts to find this ideal solution.

The present paper discusses first R. Julia's method for deriving a physically dual network by the introduction of ideal transformers. By giving the justification for this procedure more fully than it was given in his original publication we are automatically led to further methods of achieving the same result. This is useful, for some of the new methods require a smaller number of ideal transformers than Julia's original method. At the same time the fact that we are thus able to choose between a whole series of different embodiments of the dual network will be useful when we attempt to set up the dual network in a network analyser, for some of these alternative ways will possibly be less susceptible to the imperfections introduced by the use of real transformers in place of the ideal ones.

§2. CRITERIA OF DUAL CORRESPONDENCE: CONSTRUCTION OF THE DUAL COUNTERPART IN THE CASE OF PLANAR NETWORKS

The existence of the dual relationship between two networks as specified by equations (1) and (2) can be predicted very simply if we pause for a moment and consider what data determine the behaviour of a given network. There are:—

(*a*) The relations between the currents through and the voltages across the various elements of the network, i.e. a series of equations of the type

$$i_\kappa' = f(e_\kappa'); \qquad \ldots\ldots(3)$$

(*b*) The way in which these elements are connected, i.e. the relations of constraint which exist between the currents meeting at a given junction point and the voltages which we encounter when we pass around a closed mesh (Kirchhoff's equations);

(*c*) The initial state of the network (currents through and voltages across certain elements);

(*d*) The external " causes " (voltages or currents) impressed on the network.

All these data together give us a system of equations—not necessarily a linear one—the solution of which describes the behaviour of the network.

Suppose now that we have a second network for which we can list a corresponding set of data, the correspondence being of the type described by equations (1) and (2). This means that we get a corresponding set of data which is of the same form as the original one but with the difference that the symbols *i* and *e* have changed places. It is obvious then that the resultant system of equations will show the same interchange of symbols, and the same holds for the final solution of this system. Hence the second network must show a behaviour dual to that of the first network.

Let us now discuss in detail what is involved in having such an interchanged set of data between two networks. In order not to interrupt the exposition we shall deal first with the items (*c*) and (*d*). The dual counterpart of the initial conditions is self-evident; any statement of this type can be translated in a straightforward manner by the application of equations (1) and (2). The same applies to the specification of the external " sources " which are connected to the network. A voltage source corresponds to a current source of proportional strength and vice versa.

As regards (*a*), equation (3) is, according to equations (1) and (2), replaced by an equation of the type

$$\frac{e_\kappa''}{R_0^1} = f(i_\kappa'' R_1^0), \qquad \ldots\ldots(4)$$

i.e., apart from factors of proportionality, the graph of current against voltage for each element of the original system will represent the graph of volts against current in the new system. If we limit ourselves to linear systems and study the steady-state behaviour of such systems at a frequency ω we have for any particular element of the first system and for its counterpart the relations

$$e_\kappa' = i_\kappa' Z_\kappa' \quad \text{and} \quad e_\kappa'' = i_\kappa'' Z_\kappa'', \qquad \ldots\ldots(5)$$

where Z_κ' and Z_κ'' denote the impedances of the elements concerned. Multiplying these two equations and taking into account equations (1) and (2), we obtain

$$R_0^1 R_1^0 = Z_\kappa' Z_\kappa''. \qquad \ldots\ldots(6)$$

Equivalent expressions for the last equation are

$$\left.\begin{aligned} Y_\kappa' &= Z_\kappa'' \frac{1}{r_0^2}, \\ Y_\kappa'' &= Z_\kappa' \frac{1}{r_0^2}, \end{aligned}\right\} \qquad \ldots\ldots(7)$$

where $Y_\kappa' = 1/Z_\kappa'$ and $Y_\kappa'' = 1/Z_\kappa''$,

and where $r_0^2 = R_0'R_1^0$. ......(8)

From these equations it follows that the dual counterpart of a resistance R is a conductance $G = R/r_0^2$, that the counterpart of an inductance L is a capacitance $C = L/r_0^2$ and that of a capacitance C an inductance $L = Cr_0^2$.

It will be noted that these relations are independent of the frequency ω, i.e., that dual relationship once found for certain frequencies will be valid for all frequencies. On account of Fourier's theorem this means that once the dual relationship is established for steady-state conditions at one particular frequency it will hold good not only for steady-state operation at any other frequency but also for transient behaviour.

Turning now to point (*b*), the problem of interconnection of elements, we will discuss a special case first, namely the case where two or more elements in the first system are connected in series. This means that all these elements carry the same current, and that their voltages add. Their dual counterparts must then be so connected that across each of them appears the same voltage, and that all their currents add, i.e. they must be connected in parallel. This relation works, of course, the other way round. It is a very useful relation, for it can be applied for checking even when we have to deal with a more general type of network; we can always give to some of the impedances of a network the values zero or infinity, and thereby derive from the given network a number of simpler networks to which these conversion rules can be applied. (We might establish here a principle of " detailed reciprocation ", the name being chosen in analogy to the " principle of detailed balancing " as used in thermodynamics).

A more general way of establishing relations of constraint dual to the given ones is based on a consideration of Kirchhoff's equations. Suppose we have for a junction point (σ) of the given network written down the condition that the sum of all branch currents reaching this point is zero, i.e.

$$i'_{\sigma 1} + i'_{\sigma 2} + \ldots . = 0. \qquad \ldots\ldots(9)$$

It follows then that in the dual network the voltages across the counterparts of all these branches must add up to zero:

$$e''_{\sigma 1} + e''_{\sigma 2} + \ldots . = 0, \qquad \ldots\ldots(10)$$

i.e., we would normally expect these counterparts to be so arranged that they formed a closed mesh. Conversely we see that elements which are so arranged that they form a closed mesh should have dual counterparts which are joined to a common junction. We see that dual conditions of constraint are automatically introduced if the second network is in this particular geometrical relation to the first one, i.e. if it is the topological dual of the first network.

There exists a routine method for drawing such a topological dual (Cauer, 1934).

In the interior of each mesh of the given network we select a point. These points will form the junction points of the new network; we need only join each of these points to its neighbours (i.e. to those new points situated in adjoining meshes) in order to complete the layout of the dual network. Each new branch crosses one old branch (the branch which was common to the two adjoining meshes). The procedure is illustrated in figures 1 *a* and 1 *b*. In figure 1 *a* the new circuit is sketched in dotted lines, in figure 1 *b* it is redrawn, and at the same time the proper elements inserted, taking into account the relations (6) and (7) and their consequences as to the interchange of capacities and inductances. It is seen that in a case where the original branch is composed of several elements in series (e.g. 4 and 6) we could draw such a crossing line through each of these elements; all these new branches are then in parallel, and if we place into each of these new branches an element with an admittance proportional to the impedance of the "crossed" old element the resulting compound admittance will obviously be the correct dual of the original compound impedance.*

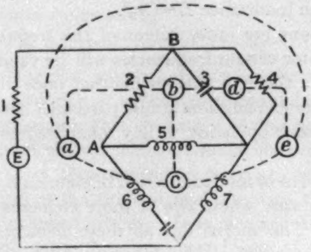


Figure 1 *a*.

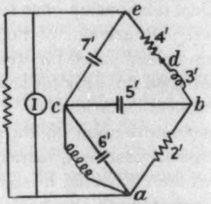


Figure 1 *b*.

The procedure just described breaks down if the original circuit diagram cannot be drawn without crossings between some of the lines ("non-planar" networks). Such networks have no topological duals, and in their case we have to use other means for introducing the required constraints between the various elements of the dual network.

§ 3. IDEAL TRANSFORMERS

We shall now enter into a discussion of these other means which will enable us to introduce the required constraints. Suppose we have an iron core with several windings of the same number of turns, i.e. a multi-winding transformer, schematically represented in figure 2 *a*. Let us assume further that the permeability of the core material is infinite and that the windings have negligible leakage. In this case we are then assured that the same flux is linked with each of the windings, i.e. that the terminal voltages of all the windings are equal. Furthermore, the finite flux required for any such finite voltage can then be produced by a magnetomotive force which is infinitesimally small, i.e. we must have

---

* It will be noted that this method of drawing the dual circuit can also be used to define in a consistent manner the sense of the dual currents or voltages. We need only agree that the arrows defining this sense in the new branches must be produced from those of the original network by, say, a rotation in a clockwise rotation.

$$i_1+i_2+i_3+\ldots\ldots=0. \qquad \ldots\ldots(11)$$

A transformer of this kind is called an *ideal transformer*. Real transformers differ from the ideal one by possessing leakage inductance which we might visualize as a separate inductance connected in series with each individual winding and by requiring a finite magnetomotive force for the creation of their magnetic flux. With real transformers the voltages across the individual windings may then no longer be exactly equal, and furthermore, equation (11) will be only approximately fulfilled. These deviations are, however, usually small, and they can possibly be allowed for if it is desired actually to build the dual network. For a discussion of principles we are certainly allowed to make use of the concept of an ideal transformer, and we see that a transformer of the type just described imposes on the currents through its windings the same constraints as if they were associated with a common junction point.

The dual counterpart to such a transformer is shown in figure 2 *b*. There the magnetic circuit is so arranged that the cores of the individual coils—which in figure 2 *a* were all connected in series—are all connected in parallel. In this magnetic circuit, the sum of all the magnetic fluxes which arrives at one of the magnetic junctions must be zero, and from this it follows that the sum of all the voltages of the windings of the various cores (which individually are proportional to the fluxes of the corresponding cores) must be zero:

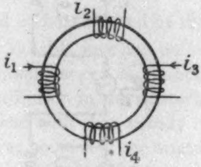


Figure 2 *a*.

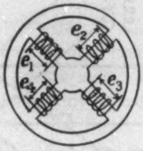


Figure 2 *b*.

$$e_1+e_2+e_3+\ \ldots=0. \qquad \ldots\ldots(12)$$

Associated with this condition is the condition that the currents through all the windings must be equal; otherwise the magnetomotive force of the individual limbs would not be equal and—in view of the assumed infinite permeability of the core material—the difference in the m.m.f. of any two of the limbs would drive a magnetic flux of infinite magnitude around the closed circuit formed by these two limbs. A transformer of the second type enforces, therefore, the same type of constraint as results from the existence of a closed mesh. Transformers with two windings are obviously special cases of these general multi-winding transformers, and we see that in their case the two types are identical.

Transformers with two windings are easier to build so as to approach the requirements of zero leakage inductance. It is therefore of interest—for those cases where it is intended to set up a physical circuit—that the same results that can be obtained by a multi-winding transformer with *n* windings can be obtained by the use of $n-1$ transformers with two windings only. If we start with n transformers with two windings we need only connect all the secondaries in series forming a closed loop, in order to enforce on all the primary voltages the condition that their sum is zero. On the other hand, if we connect all the secondaries in parallel we enforce the condition that the sum of all the primary currents is zero. As figure 3 shows (for the case $n=3$), one of those transformers is actually redundant. (For the purpose of derivation it is often convenient to retain n transformers and to carry out the reduction to $n-1$ as a last step.)

Having thus shown that these transformers will enable us to enforce constraints of the " mesh " and " junction " type,* we have to consider how the method, described in the preceding section, for the routine derivation of the dual network can be operated in the presence of such circuit elements. The dual network we would expect to be subject to transformer constraints which are the dual counterparts of the transformer constraints of the original network. It will be seen that this routine method automatically takes care of the situation provided we stipulate that each transformer coil encountered in this process is taken to have as counterpart another transformer coil assembled on an ideal transformer of the dual type. Suppose we have in the first network a number of elements $Z'_1, Z'_2, Z'_3, \ldots.$, each with such a transformer coil in series, the transformer itself being of the junction type, which insures that the sum of the currents $i'_1, i'_2, i'_3, \ldots.$ through all these elements is zero. In the new network these coils will automatically be placed in parallel to the dual counterparts $Y''_1\ Y''_2\ Y''_3$ of the elements $Z'_1, Z'_2, Z'_3 \ldots.$, and thereby enforce the condition that in the dual network the voltages $e''_1\ e''_2\ e''_3 \ldots.$ across these elements shall add up to zero. Obviously the method would have worked just as well the other way round, and would have converted a constraint of the mesh type into one of the junction type.

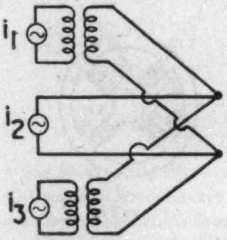


Figure 3 *a*.

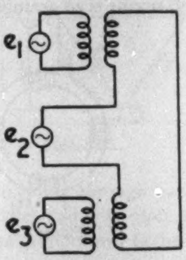


Figure 3 *b*.

It is perhaps of interest to note that the change over from the one type of transformer to its dual counterpart is actually included in the routine process of " dualization " provided we pay attention not only to the electrical but also to the magnetic circuits, the junction type of transformer having a magnetic circuit dual to that of a mesh type (figures 2 *a* and 2 *b*). If we replace the multi-winding transformer by two-winding transformers the exchange of types is simply a consequence of " dualizing " the connection of the secondary windings.

---

* We shall for this reason refer to such transformers as mesh or junction-type transformers.

§4. THE CASE OF NON-PLANAR NETWORKS

(a) *The general principle*

We are now in a position to deal with non-planar networks. Networks of this type can be converted in various ways—presently to be described—into equivalent planar networks which are subject to additional constraints by the aid of one or several ideal transformers. By ***equivalence*** we mean here that the resultant networks possess, apart from those specially added ideal transformers, the same elements as the original one, and that the currents and voltages at these elements are identical with those of the original network. This equivalent network is then converted into its dual counterpart by the routine method already described.

Of the various methods to convert a non-planar network into an equivalent planar network we shall discuss here the following—without making an attempt to be exhaustive: Julia's method (1939); a method dual to Julia's method; the method of fictitious junction points; the method of the transferred terminal; and the general method of separation.

(b) *Julia's method*

This method is characterized by the following steps:—

(1) Select a suitable junction point and introduce into each of the branches leading to this point the coil of a junction-type transformer. This will not alter the current distribution in the network as the constraint exercised by this transformer is identical with that already in existence. In figure 4 *b* the junction point selected was point A of the original network, illustrated in figure 4 *a*.

(2) Short circuit one of the branches leading to the junction point selected. This has been indicated in figure 4 *b* by the dotted line which connects A to D, and short circuits branch 1. This step again will not alter the current distribution of the network; though we have lifted by this step the constraint originally exercised by the existence of the junction point A, the transformer coils exercise now an identical constraint. Since the sum of the currents flowing through the ideal transformer windings is zero, there can be no current through the short-circuiting link.

(3) Duplicate the short-circuiting link in the way shown in figure 4 *c*, introducing thereby the auxiliary junction points A′ and D′ (figure 4 *c* shows only those parts of the original network which are essential to the argument).

(4) Cut the link AA′; this cannot alter the current distribution in the network, because the potential difference AA′ after the cut is still zero on account of the connection DD′. But now we are allowed to cut DD′ as well, for the current through this isthmus must, of course, always be zero.

By these steps we have split the original network into two entirely separate networks, the elements of which have still the same voltages across and the same currents through them as they had before. (We need not have cut DD′, for after the cut AA′ has been made it would have been possible simply to fold aside the loop containing $\mathbf{Z}_1$ and produce thereby a plane configuration.)

The new plane configuration is redrawn in figure 4 *d*. We may check that on account of the transformer action the current through element $\mathbf{Z}_1$ is still equal to the algebraic sum of the currents through elements Z_2 and Z_3; furthermore, that the voltage drop in going from D to B is equal to the voltages across elements Z_3 and Z_1, and similarly the voltage drop in going from D to F is equal to drops across Z_2 and Z_1.

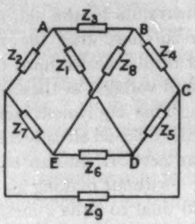

A
Z3
B
Z2
Z1
Z8
Z4
F
C
Z7
Z5
E
Z6
D
Z9

Figure 4 *a*.

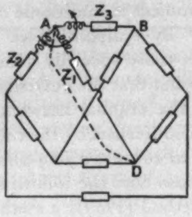


Figure 4 *b*.

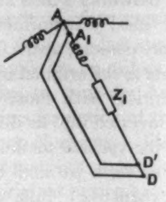


Figure 4 *c*.

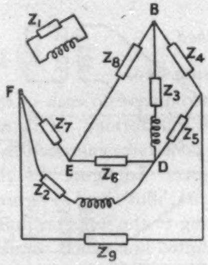

Z1
B
Z8
Z4
Z3
F
Z7
Z5
E
Z6
D
Z2
Z9

Figure 4 *d*.

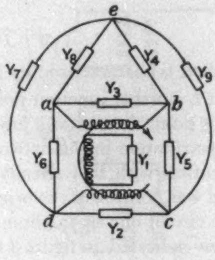

e
Y8
Y4
Y7
Y9
Y3
a
b
Y6
Y1
Y5
d
Y2
c

Figure 4 .

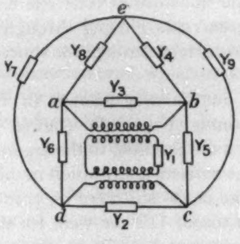

e
Y8
Y4
Y7
Y9
Y3
a
b
Y6
Y1
Y5
d
Y2
c

Figure 4 *f*.

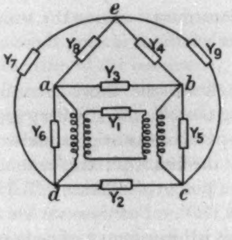

e
Y8
Y4
Y7
Y9
Y3
a
b
Y1
Y6
Y5
d
Y2
c

Figure 4 *g*.

Now the new configuration of elements, being a planar one, can be transformed into its dual counterpart by the method described above. We arrive then at the network shown in figure 4 *e*. Figure 4 *f* shows the same network with two 2-winding transformers. Figure 4 *g* is included for the sake of interest: it shows the final network at which we would have arrived had we eliminated from the original network the junction point D instead of A.

(c) *The dual counterpart to Julia's method*

We come now to the description of a new method for the derivation of an equivalent planar network which can be characterized as the dual counterpart of Julia's method. In Julia's method we used transformer action to replace a junction type of constraint. Consequently, the dual network arrived at is one in which transformer action is used to implement a mesh type of constraint.

Obviously it is possible to invert this procedure, i.e. we can start by replacing a mesh type of constraint by transformer action and finish them with a dual network in which transformers implement a junction type of constraint. To illustrate this procedure we shall carry it out with the same network which we used just now, and we shall again eliminate from this network branch 1. To achieve this we proceed in the following steps:—

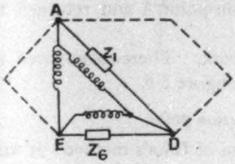


Figure 5 *a*.

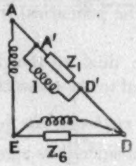


Figure 5 *b*.

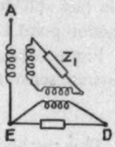


Figure 5 *c*.

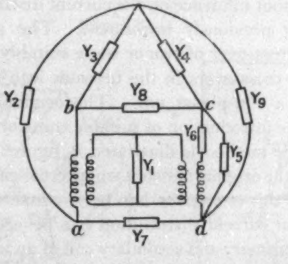


Figure 5 *d*.

(1) Select a suitable mesh which contains the branch to be eliminated. The mesh to be selected may also include such branches of the original network as contain elements of infinite impedance. Thus, for instance, we may select in figure 4 *a* the mesh ADEA, of which the branch AE is of infinite impedance, and therefore not explicitly shown as "branch" in the diagram. Place in parallel to each mesh branch the coil of a mesh type of ideal transformer (or the input coil of an ideal two-winding transformer, the secondaries of which are all connected in series). This step, illustrated in figure 5 *a*,* cannot upset the current distribution in the network, for the constraint thereby imposed is identical with one already in existence.

(2) Re-draw this network in the way shown in figure 5 *b*, introducing thereby the auxiliary junction points A′ and D′. Cut the link AA′. This further step

---

* The figure shows only those branches which are relevant to the argument.

is permissible without upsetting the network performance as, on account of the transformer action, the potential difference across coil 1 is maintained independent of any current flowing therein; hence the potential difference AA′ after the cut must still be zero.*

(3) Cut the link DD′; this is permissible now—after the cut AA′ has been carried out as the current in the isthmus DD′ must be zero.

By these steps we have again removed branch 1 from the original network. Figure 5 *c* shows how the same end can be achieved by the use of 2 two-winding transformers. Here it is immediately obvious that the two transformers supply a potential AE + ED = AD to element 1, and that the current drawn by this element is in fact still drawn (via the primaries) from junction point A and returned to junction point D.

Figure 5 *d* shows the dual to this latter network. There is no need to illustrate separately the dual to the network shown in figure 5 *b*.

(d) *The method of the fictitious junction point*

This method can be considered as a generalization of Julia's method; it will as a rule lead to a smaller number of ideal transformers.

Julia's method was characterized by the short circuiting of a branch—this short circuit being without influence on the current distribution on account of the transformer constraint previously introduced. The new method introduces a short circuit at the cross-over of two or more suitably selected branches. By converting thus all the cross-overs in the network into real junction points, the network ceases to be a non-planar one. The original current distribution is again maintained by the introduction of suitable transformer constraints.

The principle of the method is illustrated in figures 6 *a* to 6 *c*. Figure 6 *a* shows two branches of the original network which cross each other without making contact. In figure 6 *b* this cross-over has been converted into a real junction point (X); the original current distribution has nevertheless been maintained by the insertion of the primary and secondary coil of an ideal transformer into the lead BD. The action of this transformer ensures that as much current leaves the new junction point X through Z_4 as arrives there through Z_2. Hence none of this current can leave through Z_1 or Z_3. (The introduction of another transformer into the branches containing the elements Z_1 and Z_3 would of course be permissible, and may in a practical case aid in the maintenance of this state of affairs). Figure 6 *c* finally shows the dual counterpart of the network section shown in figure 6 *b*.

An application of the principle just discussed to the network of figure 4 *a* is shown in figures 7 *a* and 7 *b*. It will be noted that by this method we save—compared with the solutions previously given—one ideal transformer. This saving in a number of ideal transformers is larger still in the case of the network shown in figure 8 *a* which represents the second example given in Julia's paper. In dealing with this network Julia removes the junction point A, and requires for this purpose either an ideal transformer with four windings or three transformers with two windings. From inspection of figure 8 *a* it follows immediately

---

* A somewhat more detailed argument is as follows: As the potential across coil 1 remains unchanged, the current through element Z_1 remains also unchanged. After the cut, this current must be supplied from the transformer coil 1, and for this reason all the other coils must then carry a current equal to this current. This means that at junction point A a current equal to this current is still leaving (into coil AE), and at junction point D a current of equal magnitude is still arriving.

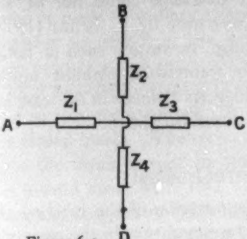

B
Z2
Z1
Z3
A
C
Z4
D

Figure 6 *a*.

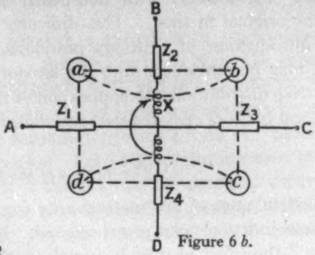

B
Z2
a
b
X
Z1
Z3
A
C
d
c
Z4
D

Figure 6 *b*.

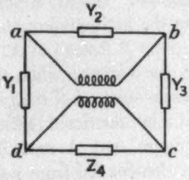

Y2
a
b
Y1
Y3
d
c
Z4

Figure 6 *c*.

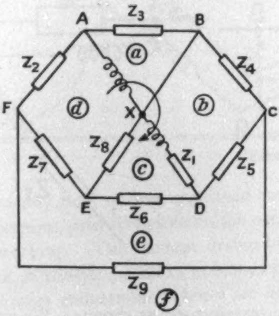

A
Z3
B
Z2
a
Z4
F
d
X
b
C
Z8
Z1
Z7
c
Z5
E
Z6
D
e
Z9
f

Figure 7 *a*.

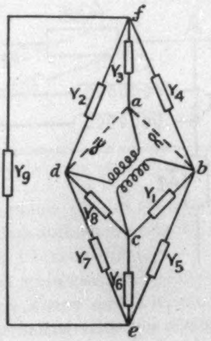

f
Y3
Y2
Y4
a
Y9
d
b
Y1
Y8
c
Y7
Y5
Y6
e

Figure 7 *b*.

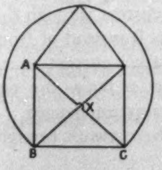


Figure 8 *a*.

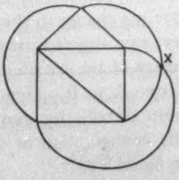


Figure 8 *c*.

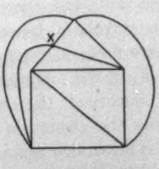


Figure 8 *b*.

that this case can be dealt with by one ideal transformer with two windings only. Figures 8 *b* and 8 *c* show alternative solutions, obviously without exhausting the numerous possibilities.

The artifice of the fictitious junction point here introduced is of some further theoretical interest. As is well known, reciprocal diagrams of the type discussed occur also in graphical statics as reciprocal stress diagrams to pin-jointed frameworks. Frameworks with non-planar lay-out diagrams could not be solved by the original method. This difficulty was overcome by F. Schur (1897) by the introduction of fictitious pin-joints. Whilst, however, such a fictitious joint may be introduced into the lay-out of the framework without upsetting the stress distribution, the application of the analogous artifices in the case of the electrical network requires the special measures discussed above.

(e) *The method of the transferred terminal*

Certain cases of this method arise from an application of the method of the fictitious junction point just discussed. It will be noted that in the example of figure 7 the effect of introducing the fictitious junction point is that now Z_1 is connected directly * to junction point B instead of A. Such a " transfer of the terminal " will always take place, for instance, when the two leads crossing each other contain only one element each—instead of the two assumed in figure 6.

However, the transfer method is not restricted to the presence of such crossing branches. Quite generally if (figure 9 *a*) A and B represent two junction points of the network, and Z_t an element fed from point A, we can join A to B by the primary and secondary of an ideal transformer, as shown. We are then allowed to break the connection between primary and secondary at X—for on account of the transformer action no difference of potential can arise between the separated leads and, therefore, no change in the current through Z_t can take place. This current is, however, now supplied from B instead of from A; in addition, there flows in the primary of the transformer a current equal to that taken from B, in the direction from A to B. Hence, as far as the rest of the network is concerned, the current distribution remains unchanged. For the sake of clarity the change in the configuration of the network brought about by this step is illustrated again in figure 9 *b*.

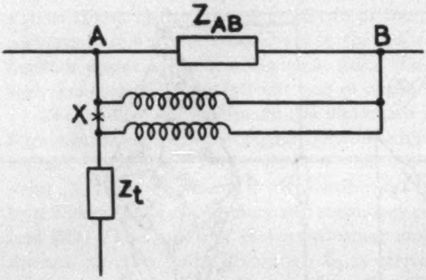


Figure 9 *a*.

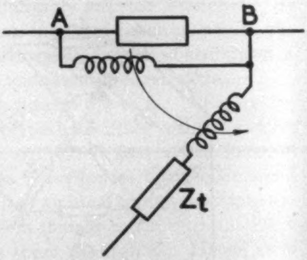


Figure 9 *b*.

---

* It is true that in the diagram shown, the transformer coil is interposed between the terminal of B and A; however, the diagram could obviously have been drawn with the order of these two elements in branch 1 inverted.

(f) *The general method of separation*

This method, which can be described as a generalization of the dual to Julia's method, is illustrated in figure 10. Instead of eliminating a single branch from the original network we eliminate at one time several branches which together form one or more subsidiary networks.

Suppose we select in the original network a closed mesh containing the junction points (in this order) MNPQ . . . . and we join M to N, by the primary of an ideal two-winding transformer, similarly N to P, and so on. We are then allowed to join consecutively all the secondaries of these transformers so as to form a closed mesh. The mesh of the secondaries forms a kind of " parallel " mesh to the mesh formed by the primaries. Let us label the junction points of this second mesh M′N′P′Q′ . . . . It is obvious that the potential difference between any one pair of these latter points is a replica of the potential difference of the corresponding unprimed points. For this reason any branch extending between any pair of the unprimed junction points can be transferred to the corresponding primed points on the parallel mesh without altering the conditions of its working. The current drawn by such a branch still leaves the original network at the old junction points as currents which feed the primaries of the transformers concerned. Hence we have here a very simple means by which we can split the original network into two subsidiary networks which are only coupled by transformer action.

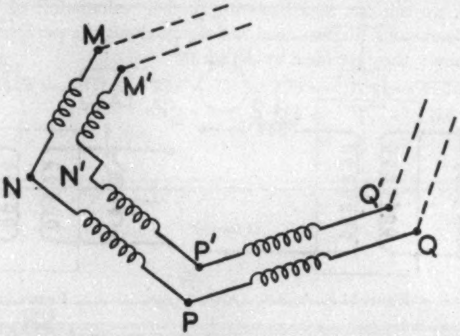


Figure 10.

It is easily seen that the transformers need not actually form a closed mesh: one of the sides of the mesh may be left out. Furthermore, the mesh selected may—as was the case before (figure 5 *a*)—contain branches of infinite impedance.

This concludes the description of the various methods of network transformation which, it is felt, covers all needs arising in practice. No attempt has been made to cover the subject exhaustively.

Application

The possibility of constructing without restriction a network that is physically dual to a given network is in itself only of academic interest. It may, however, become of immediate practical interest in connection with the increasing use of electro-mechanical analogies. As shown elsewhere (e.g. Bloch, 1945), we can study the action of an electro-mechanical system if we connect the electrical part of this system to a network which is the electrical analogue of the mechanical part of this system; this reduces the analysis to that of a single electrical network. However, this possibility exists only when we use the correct type of analogy: the direct one for the electrostatic type of coupling element and the inverse one for the electromagnetic type of coupling element. For the reasons already mentioned in the introduction to this paper, this method would break down—but for the methods here described—if we had to use the direct analogy in a case where the mechanical layout, and therefore the inverse analogue, were non-planar.

The methods of this paper enable us

(*a*) To draw a diagram of the direct analogue, and hence of the combined network, which we can use as a basis for further calculation or consideration. In this case it is obvious that we need not worry about the performance of those ideal transformers; they are ideal by definition.

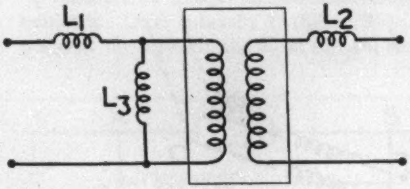


Figure 11 *a*.

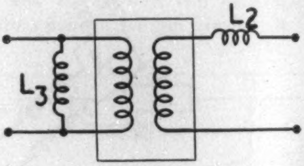


Figure 11 *b*.

(*b*) To set up an actual circuit, say, on a network analyser which will enable us to study directly by measurement the performance of the electrical circuit or of the underlying mechanical system. In this case the ideal transformers have to be replaced by real ones. As real transformers do not behave exactly like ideal ones, errors may arise from this, and it is here that the possibility of choosing from a large number of different methods becomes of value; it may be possible then to select a method in which the imperfection of the real transformers are innocuous.

Figure 11 (*a*) shows a well-known equivalent circuit for a real transformer, i.e. a theoretical conception which explains the behaviour of a real transformer as arising from the combination of an ideal transformer with certain inductances: leakage inductances L_1 and L_2, and shunt inductances L_3. It is perhaps worth while to point out that the distribution of the leakage inductance on the primary and secondary side of the transformer is to some extent arbitrary, and depends on the assumption made as to the magnitude of the ratio of the ideal transformer. By a suitable assumption it is possible to have all the leakage inductance concentrated on the primary or on the secondary side. We arrive thus at an equivalent circuit such as shown in figure 11 *b*.

The error caused by the leakage inductance is of course small in all those cases where the impedance of this additional inductance is small compared with all the other impedances in series with it. Similarly the error caused by the presence of the shunt inductance will be small when its admittance is small compared with the other admittances across which it is connected. In certain cases it will be possible to reduce these errors actually to zero—at least as far as the rest of the network is concerned. Suppose the transformer coil is connected in series with another inductance larger than the leakage inductance. Obviously we need then only reduce this other inductance by the amount of the leakage inductance in order to arrive at a combination which to the rest of the network looks exactly like the ideal combination. Similarly we can deal with the case of a shunt inductance in parallel to another inductance. The equivalent circuit of figure 11 *b*, in which all the leakage inductance is concentrated on one side, is specially useful for dealing with such cases. The combinations just referred to are wanted in the physical dual circuit; this means that when planning the original equivalent planar network we have to look ahead for their dual counterparts. Hence, for instance, when we wish to eliminate entirely the error caused by the leakage inductance we have to look for an equivalent network in which one transformer coil is placed in parallel to a capacity of sufficient magnitude.

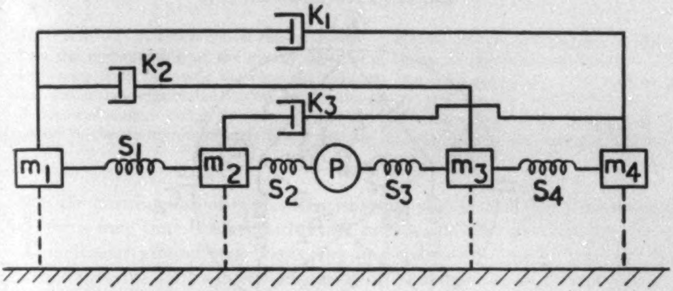


Figure 12.

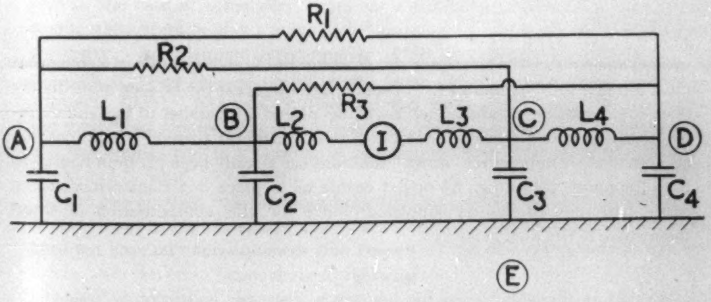


Figure 13.

Similarly, we have to place one transformer coil in series with a capacitative impedance of sufficient magnitude when we finally wish to eliminate the error caused by the shunt inductance.

An application of the aforesaid principles and of the principles of this paper in general is illustrated in figures 12 to 14. Figure 12 depicts a mechanical system consisting of a prime mover and of four masses which are inter-linked by an arrangement of springs and viscous couplings. The mode of presentation follows the convention which facilitates the drawing of the circuit of the electrical inverse analogue. This circuit is shown in figure 13, and redrawn in a somewhat more familiar form in figure 14. From the latter figure it can be seen immediately that, for instance, a transformer could be placed in parallel to C_2, and thereby the terminal of C_1 transferred from E to B. Other "parallel" solutions are also obvious. A number of "series" solutions has already been given in figures 8*a* to 8*c*; some of these (e.g., 8*b* or 8*c*) enable us to place one transformer coil in series with a capacity—as stipulated above for the compensation of shunt inductance in the dual network.

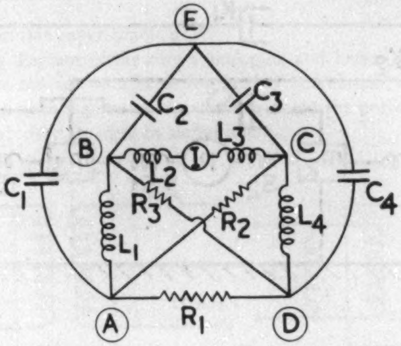


Figure 14.

REFERENCES

BARNES and GARDENER, 1942. *Transients in Linear Systems* (John Wiley), p. 49.

BLOCH, A., 1945. *J. Inst. Electr. Engrs.*, **92**, Part I, 157.

CAUER, W., 1934. *Z. angewandte Math. Mech.*, **14**, 349.

JULIA, R., 1939. *Bull. Soc. Franç. Electr.*, **9**, 28.

RUSSELL, A., 1904. *Alternating Currents*, chapter XVII (Cambridge University Press).

SCHUR, F., 1897. *Math. Annal.*, **48**, 142.

SIRE DE VILAR, H., 1901. *L'Eclairage Electrique*, **27**, 252.

WHITNEY, HASSLER, 1931. *Proc. Nat. Acad. Sci.*, **17**, 125.

ELECTRONIC STATES AND POTENTIAL CURVES OF DIATOMIC MOLECULES

By J. G. VALATIN,
Budapest

Communicated by W. Jevons. MS. received 28 *March* 1946 ;
in revised form 4 *June* 1946

ABSTRACT. A coordination rule of discrete molecular energy terms cannot be deduced from the non-crossing of the energy eigenvalue curves of the two-centre model. Clear distinction is to be made between the potential curves of unperturbed electronic states and the energy eigenvalue curves of the two-centre problem, the approximate existence of potential curves being the consequence of the averageability of the interactions with respect to the electronic motions which can also be influenced by the motion of the nuclei.

In the coordination of molecular terms with atomic term combinations, it has for a long time been regarded as a quantum-mechanical principle that the molecular ground state dissociates into the atomic ground states, if these atomic states can form a molecular state of the corresponding symmetry (Herzberg, 1937; Hulthén, 1939). As has been pointed out several times by Schmid and Gerö, this coordination rule is inconsistent with strong spectroscopic arguments even in the case of some very important diatomic molecules, and the excited atomic states seem to play a much more important rôle from the point of view of chemistry. An adequate interpretation of all the spectral data leads to dissociation schemes for CO (Schmid and Gerö, 1937), CN (Schmid, Gerö and Zemplén, 1938), NO and N_2 (Schmid and Gerö, 1943), according to which the molecular ground state dissociates into higher atomic term combinations, and this applies also to a number of diatomic hydride and deuteride molecules (Gerö and Schmid, 1939, 1940).

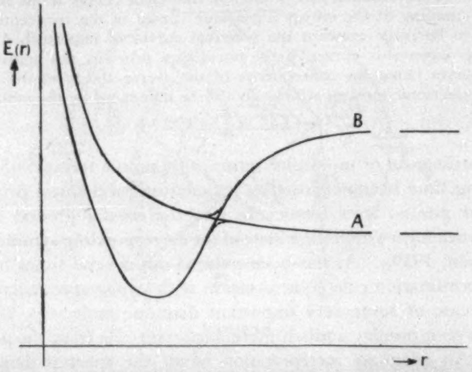
The coordination rule mentioned resulted from a direct adaptation of the conclusions concerning the energy curves of the two-centre problem, whereas these need not have any consequences with respect to the discrete molecular energy levels. On the other hand, sharp distinction is to be made between the potential curves of unperturbed electronic states and the energy eigenvalue curves of the two-centre problem, the connection being much less close than has often been supposed. The motion of the electrons can also be influenced by the motion of the nuclei, and the wave-function factor characterizing an unperturbed electronic state is not identifiable with the eigenfunction of the two-centre problem in general.

We write the wave equation of the two-centre problem in the form

$$\{T_e + U\}\Phi_0(x_i, r) = E(r)\Phi_0(x_i, r), \qquad \ldots\ldots(1)$$

where x_i represents the coordinates of all the electrons, T_e symbolizes the kinetic energy of the electrons, U the interactions between all the particles; the fixed internuclear distance r occurs only as a parameter; and the energy eigenvalues

E and eigenfunctions Φ_0 of the problem result as functions of this parameter. As shown by Neumann and Wigner (1929), if certain symmetry conditions are fulfilled, two energy eigenvalues differing at some value of the continuously variable parameter r cannot coincide at another value of the internuclear distance, and energy curves of the same species do not cross. If in a certain approximation (for instance, in the approximation of the Heitler-London theory) two energy curves cross, then, taking into account the neglected terms, the energy values belonging to fixed values of r repel each other, and the curves avoid crossing. The figure shows how, with respect to the energy eigenvalue curves of the two-centre problem, this leads to the coordination rule mentioned. The corresponding conclusions have been applied also to the coordination of molecular and atomic terms (Hund, 1927).



From the point of view of the molecule showing discrete energy levels, however, the nuclear distance cannot be regarded as a continuously variable parameter. The energy operator of the molecule, containing also a differential operator with respect to r, does not commute with the internuclear distance; energy and internuclear distance of the molecule are not even simultaneously measurable quantities in quantum mechanics. At a given value of the energy only the probability distribution of the r values is defined by the radial wave-function factor of the corresponding state. There is no sense in speaking of the dependence of energy on the nuclear distance, and the energy states of a molecule cannot even approximately be related to definite values of r. Accordingly, the corresponding conclusions cannot be transferred in this way from this model to the sequence of discrete molecular terms.

In the usual discussion of the diatomic molecule (Van Vleck, 1929; Kronig, 1930) it is supposed that the wave equation of the molecule can be separated into three parts, the neglected terms causing the perturbations observed in molecular spectra. One of these equations is the wave equation of the two-centre problem, the energy eigenvalue function of which enters into the equation describing nuclear vibrations, the potential curve of the molecule being defined in this way by the energy eigenvalues resulting at fixed nuclei. The potential function of an unperturbed electronic state, however, can also be interpreted independently from these energy curves as the force law determining the nuclear motion which results from an averaging of all the interactions with respect to the electronic motions. This separation means a rather arbitrary division of the coupling between electronic and nuclear motions, which, in general, cannot be justified mathematically.

Using relative coordinates of the particles related to one of the nuclei, and using a reference system bound to the nuclei, the axes of which are given with respect to an inertia system by the Eulerian angles ϑ, ψ, ϕ, its z-axis showing the direction of the line joining the nuclei, we can write for the wave equation of a diatomic molecule,

$$\{T_r + T_e + U\}\Psi(x_i, r, \vartheta, \psi, \phi) = E\Psi(x_i, r, \vartheta, \psi, \phi), \qquad \ldots\ldots(2)$$

neglecting correction terms which can be ascribed to the finite mass connected with the origin of reference system. T_r now represents the relative kinetic energy of the nuclei, which involves the reduced mass, $\mu = M_1M_2/(M_1+M_2)$, of the nuclei. We can restrict ourselves to singlet states, the inclusion of the spin in the usual way involving no complications.

Simultaneously with the energy equation, in the chosen reference system with respect to the total angular momentum of the molecule, we have

$$M^2\Psi = -\left\{\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\left(\sin\vartheta\frac{\partial}{\partial\vartheta}\right) + \frac{1}{\sin^2\vartheta}\left(\frac{\partial}{\partial\psi} - iL_z\cos\vartheta\right)^2 - L_z^2\right\}\Psi = J(J+1)\Psi, \qquad \ldots\ldots(3)$$

L_z denoting the operator corresponding to the component of the resultant electronic angular momentum in the direction of the nuclear axis. The rigorous validity of this equation, giving eigenvalues of a constant of motion, simultaneously with the wave equation, is to be especially emphasized, since this equation multiplied by the factor $\frac{h^2}{8\pi^2\mu}\frac{1}{r^2}$ often appears in molecular literature as the wave equation with only an approximate character of the rigid symmetrical top (Van Vleck, 1929; Kronig, 1930).

Denoting the corresponding components of the total and the resultant electronic angular momenta by M_x, M_y and by L_x, L_y, we can in first approximation neglect terms in the wave equation which represent a coupling between rotation and electronic motion, and which can be written in the chosen reference system as

$$-\frac{h^2}{4\pi^2\mu}\frac{1}{r^2}(M_xL_x + M_yL_y)\Psi; \qquad \ldots\ldots(4)$$

and we can take wave functions Ψ_Ω satisfying the restriction

$$-i\frac{\partial}{\partial\phi}\Psi_\Omega = L_z\Psi_\Omega = \Omega\Psi_\Omega. \qquad \ldots\ldots(5)$$

These can be factorized according to

$$\Psi_\Omega(x_i, r, \vartheta, \psi, \phi) = \Phi_\Omega(x_i, r)\,\Theta_\Omega^{J,M}(\vartheta, \psi, \phi). \qquad \ldots\ldots(6)$$

The function $\Theta_\Omega^{J,M}(\vartheta, \psi, \phi) = e^{iM\psi}e^{i\Omega\phi}Q_\Omega^{J,M}(\cos\vartheta)$ is a solution of equation (3), and we get for the determination of $\phi_\Omega(x_i, r)$ the equation

$$\left\{\frac{h^2}{8\pi^2\mu}\left(-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{L_x^2+L_y^2-L_z^2}{r^2}+\frac{J(J+1)}{r^2}\right)+T_e+U\right\}\Phi_\Omega(x_i,r)$$
$$=E\Phi_\Omega(x_i,r). \qquad \ldots\ldots(7)$$

In accordance with the rigorous character of the rotational quantum number J, we have a natural way of separating the angular variables. The effect of the rotational coupling can be followed in the function of the quantum number J, the effect of the terms under expression (4) being negligible for $J \to 0$. We have another situation with respect to the coupling between nuclear vibration and electronic motion, which also is fully contained in equation (7). According to the purely dynamical character of nuclear vibration, there is no quantum number which retains its rigorous significance even in the case of strong coupling effects, and every approximate separation can by its nature have only a dynamical justification.

Supposing, with respect to the wave functions of an unperturbed electronic state,

$$\Phi_\Omega(x_i, r)=f(r)\,\Phi_e^{(\Omega)}(x_i, r) \qquad \ldots\ldots(8)$$

with

$$\int\overline{\Phi}_e^{(\Omega)}\Phi_e^{(\Omega)}dx_i\equiv 1, \qquad \ldots\ldots(8a)$$

we can multiply equation (7) by $\overline{\Phi}_e^{(\Omega)}(x_i, r)$ and integrate on the electronic coordinates. For the wave functions $f(r)$ characterizing nuclear vibration we get in this way the wave equation

$$\left\{\frac{h^2}{8\pi^2\mu}\left(-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{J(J+1)}{r^2}\right)+V(r)-E\right\}f(r)=0, \qquad \ldots\ldots(9)$$

which yields the regular term system of an unperturbed electronic state of the molecule, with a potential function

$$V(r)=\int\overline{\Phi}_e^{(\Omega)}(T_e+U)\,\Phi_e^{(\Omega)}dx_i+\frac{h^2}{8\pi^2\mu}\int\overline{\Phi}_e^{(\Omega)}\left(-\frac{\partial^2}{\partial r^2}+\frac{L_x^2+L_y^2-L_z^2}{r^2}\right)\Phi_e^{(\Omega)}dx_i, \qquad \ldots\ldots(10)$$

resulting from averaging the interactions.

The nuclear vibration is determined in this way by the motion of the electrons through the interaction of the particles; but, owing to the same interactions, we have to suppose that the motion of the nuclei can also influence the motion of the electrons with respect to which the averaging is performed. There is no mathematical reason to suppose that the functions $\Phi_e^{(\Omega)}$ can be identified with the eigenfunctions Φ_0 of the two-centre problem in general, and the first term in the form (10) of the potential function is identifiable with an energy eigenvalue function $E(r)$ of the two-centre problem. Wave equation (2) contains symmetrically the kinetic energies T_r of the nuclei and T_e of the electrons, and it is not justifiable to neglect the motion of the nuclei simply in determining the electronic motion, this involving a generally arbitrary division of the coupling between nuclear vibration and electronic motion. The corresponding hypothesis would lead also to the experimental difficulties mentioned, and is not warranted by theory.

Potential curves can be defined only for the unperturbed electronic states. In passing from equation (7) to equation (9) we replaced the corresponding energy terms by their averages, neglecting coupling terms, to the effect of which can also be ascribed the homogeneous perturbations observable between the discrete molecular energy levels. The size of the displacements in the regular sequence of the levels is determined also in the case of homogeneous perturbations primarily by the actual position of the perturbing levels at a given value of J; while, in the interpretation of the effect, an important rôle should be ascribed also to the non-diagonal matrix elements of T_e+U—a point which seems to be left out of consideration in the interpretation of homogeneous perturbations by identifying the functions $\Phi_e^{(\Omega)}$ and Φ_0.

The solution of the molecular wave equation with the aid of the energy curves of the two-centre problem takes its origin from a paper of Slater (1927), where an analogous process is applied to the computation of the energy values of the helium atom, neglecting first the kinetic energy of one of the electrons and dealing with its coordinates as parameters. In the case of the much more complex problem of the molecule, this formal process may or may not give a correct approximation, and also the complications connected with an energy continuum are involved; for instance, the crossing curves of the figure as potential curves result in one discrete energy set, whereas the uncrossed energy curves of the two-centre problem would result in two energy sets.

In connection with the identification of the potential curves with the energy eigenvalue curves of the two-centre problem generally, a paper of Born and Oppenheimer (1927) is quoted, the citations often going farther, however, than the conclusions which can be drawn from this paper. In the second-order terms for the nuclear vibration a formal expansion process leads to the equation of a harmonic oscillator containing the curvature of the corresponding energy eigenvalue curve at the place of its minimum, but this does not involve any consequences with respect to an identification in the above sense.

Another formal solution with respect to the wave functions of the electronic states could be given by the composition and antisymmetrization of the wave functions of two atomic states. This also would necessarily contain the variable r and would still be multiplied by a function of r according to the normalization condition (8 *a*). Naturally, this formal approximation would not be suitable in general, but there may be cases when the potential curves resulting in this way give a still better approximation concerning the whole term system of the electronic state than the exact eigenvalue curves of the two-centre problem.

In the figure, if the crossing energy curves A and B represent the approximations of the Heitler-London theory, a perturbation calculation between the two energy curves leads in the two-centre problem to the non-crossing energy curves with wave functions

$$\Phi_0(x_i, r) = c_1(r)\,\Phi_0{}^{A}(x_i, r) + c_2(r)\,\Phi_0{}^{B}(x_i, r), \qquad \ldots\ldots(11)$$

where the factors $c_1(r)$, $c_2(r)$ may suffer a rapid variation between the values 0 and 1 near to the point of intersection. The approximations made by passing from equation (7) to equation (9) are also connected with the derivatives of $\Phi_e^{(\Omega)}(x_i, r)$ with respect to r; hence, in identifying these solutions $\Phi_0(x_i, r)$ of the two-centre problem with the wave function of the electronic state, the terms neglected may become very significant, and we cannot even be sure whether the further approximations made between the resulting discrete energy levels will give converging results at all. But, in the case of a too rapid variation of the corresponding function with r, we by no means get a good approximation of the energy values which could characterize an unperturbed term system of the molecule.

In a paper dealing with atomic collision processes, London (1932) arrives at " nonadiabatic " transitions just from the rapid variation of the factors $c(r)$ in function (11), following the corresponding calculations also in detail. Naturally, we cannot identify ourselves with the stand-point of this paper (particularly from the point of view of the discrete energy levels), which starts from the wave functions corresponding to the energy states of the two-centre problem, and ascribes physical significance first of all to these. A discrete energy level of the molecule may be determined by a potential curve which differs essentially from the energy eigenvalue curve of the " adiabatic " problem, and may show predissociation effects (which are of importance in chemical reactions), the transition probability from the stable state to the continuum being connected with the matrix elements of $T_e + U$ as well as with those of T_r.

The possibility of a " potential theoretical treatment " of some molecular problem, the investigation of a one-body problem with a potential curve, is not restricted to the model of fixed nuclei, but arises from the averageability of the interactions with respect to the electronic motions in consequence of the smaller mass of the electrons. These electronic motions may also be influenced by the nuclear motion, and the potential curves cannot be identified in general with the eigenvalue curves of the two-centre problem. Especially, it is by no means possible to infer from the behaviour of the energy curves of this model a coordination rule of the molecular energy levels, or to consider the corresponding coordination rule as supported by theoretical arguments.

ACKNOWLEDGMENTS

I dedicate this little paper with the feeling of deepest gratitude to my dear teacher, R. F. Schmid, who regarded the elucidation of the problem as his own, who followed the development of the paper with much helpful interest, and who, at the time of publication, is no longer of this world. The late Dr. L. Gerö gave me the benefit of much valuable discussion and advice.

REFERENCES

Born, M. and Oppenheimer, R., 1927. *Ann. Phys., Lpz.* (4), **84**, 457.
Gerö, L. and Schmid, R. F., 1939. *Z. Phys.*, **111**, 588 ; 1940. *Ibid.*, **115**, 57.
Herzberg, G., 1937. *Chem. Rev.*, **20**, 145.
Hulthén, E., 1939. *Z. Phys.*, **113**, 126.
Hund, F., 1927. *Z. Phys.*, **42**, 100.
Kronig, R. de L., 1930. *Band Spectra and Molecular Structure* (Cambridge University Press).
London, F., 1932. *Z. Phys.*, **74**, 143.
Neumann, J. v. and Wigner, E., 1929. *Phys. Z.*, **30**, 467.
Schmid, R. F. and Gerö, L., 1937. *Z. phys. Chem.* (B), **36**, 105.
Schmid, R. F., Gerö, L. and Zemplén, J., 1938. *Proc. Phys. Soc.*, **50**, 283.
Schmid, R. F. and Gerö, L., 1943. *Csillagászati Lapok*, **6**, 101, 102.
Slater, J. C., 1927. *Proc. Nat. Acad. Sci.*, **13**, 423.
Van Vleck, J. H., 1929. *Phys. Rev.*, **33**, 467.

PHOTOCHEMICAL DECOMPOSITION OF CO

By R. F. SCHMID and L. GERÖ,
Budapest [*]

MS. received from J. G. Valatin and communicated by W. Jevons 18 *July* 1946

The photochemical decomposition of the CO molecule was investigated by Faltings, Groth and Harteck (1938). These authors irradiated CO gas at atmospheric pressure with a xenon lamp, the radiation consisting chiefly of the 1470 and 1295 A. resonance lines of xenon. In a path of 30 mm. length the CO gas absorbed about 70% of the incident light. Only the shorter wave-length was effective, causing CO_2 formation with a quantum efficiency of about unity; the 1470 A. line does not give rise to any CO_2 formation. Decomposition of CO was presumed to be the primary process; from this it was concluded that 219 kcal. (corresponding to the shorter wave-length) must be an upper limit and 193 kcal. (corresponding to the longer wave-length) probably a lower limit for the dissociation energy of the CO molecule.

As Gaydon and Penney (1942) have pointed out, no continuous absorption is known in the CO spectrum in the neighbourhood of the energy level of the 1295 A. xenon line; only the discrete bands of the Fourth Positive CO system lie in this part of the spectrum, and the *P*(35) line of the (10,0) band coincides roughly with the xenon line. But, as the excitation of such a high rotational term of the ground state at room temperature is very improbable, they conclude that there is an absorption in the (13,2) band of the Fourth Positive system. If this is so, the primary process would then be:

$$\mathrm{CO}(\mathrm{X}\,^1\Sigma)+h\nu=\mathrm{CO}(\mathrm{A}\,^1\Pi),$$

$\mathrm{A}\,^1\Pi$ being the upper state of the Fourth Positive band system. The mean life-time of the CO molecule in the discrete $\mathrm{A}\,^1\Pi$ terms cannot be much less than about 10^{-8} sec. In this way at least a hundred collisions take place before a quantum of fluorescence radiation could be emitted at room temperature and atmospheric pressure; these collisions with unexcited CO molecules would give rise to the formation of CO_2:

$$\mathrm{CO}(\mathrm{A}\,^1\Pi)+\mathrm{CO}(\mathrm{X}\,^1\Sigma)\rightarrow \mathrm{CO_2}+\mathrm{C}.$$

No conclusion in respect of the CO dissociation energy is available in this case.

Now the above assumption of Gaydon and Penney concerning the primary photochemical process is quite impossible. As the lines of the $\mathrm{A}\,^1\Pi-\mathrm{X}\,^1\Sigma$ band system of CO are rather sharp, the half-value width of the upper-state rotational terms cannot be greater than, say, $0{\cdot}1\ \mathrm{cm.}^{-1}$; with at least this accuracy the wave-number of a CO band line should then coincide with the wave-number of the absorbed xenon line. The extrapolated wave-numbers of the CO $\mathrm{A}\,^1\Pi-\mathrm{X}\,^1\Sigma$ (10,0) band lines in the neighbourhood of the xenon line are (Gerö, 1936; Schmid and Gerö, 1936; Gerö *et al.*, 1927):

$P(35)=77213{\cdot}6\ \mathrm{cm.}^{-1}$, $Q(37)=77227{\cdot}0\ \mathrm{cm.}^{-1}$, $R(40)=77207{\cdot}1\ \mathrm{cm.}^{-1}$,
$P(36)=77170{\cdot}9\ \mathrm{cm.}^{-1}$, $Q(38)=77184{\cdot}5\ \mathrm{cm.}^{-1}$, $R(41)=77164{\cdot}1\ \mathrm{cm.}^{-1}$.
$Q(39)=77141{\cdot}1\ \mathrm{cm.}^{-1}$,

---

[*] It is much regretted that both authors have died since the preparation of the first draft of this paper.

The wave number of the xenon line is 77185·9 cm.$^{-1}$, calculated from the term values of Bacher and Goudsmit (1932); this wave number is nearest to the wave number of the $Q(38)$ line, but the accuracy of the coincidence is by no means sufficient.

The rotational quantum number belonging to the lower state of the $Q(38)$ line is $J=38$. The probability of a state with such a high rotational quantum number is very low. Figure 1 contains the percentage probabilities of the rotational terms of the CO $\mathrm{X}\,^1\Sigma\,(v=0)$ state at 300° K., evaluated according to the Boltzmann law:

$$100\,\frac{(2J+1)e^{-\frac{hc}{kT}BJ(J+1)}}{\sum\limits_{J}(2J+1)e^{-\frac{hc}{kT}BJ(J+1)}}.$$

The maximum probability lies at $J=7$; at $J=38$ its value is less than 0·0001%. The probability of the $\mathrm{X}\,^1\Sigma(v=2)$ state is even less; besides, no lines of a (13,2) (extrapolated) band lie in the spectral region of the 1295 A. xenon line, the head of this red-degraded band being at 77059 cm.$^{-1}$.

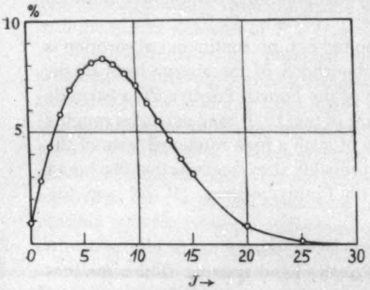


Figure 1.

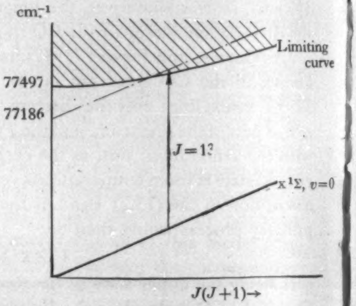


Figure 2.

But, furthermore, a third reason makes the absorption of the xenon line by a single band line quite impossible. Harteck and Oppenheimer (1942) have observed that the lines in the radiation of the xenon lamp are strongly broadened, on account of the great intensity of the current used. Now these broad lines cannot be absorbed up to 70% by discrete band lines, but only by a region of continuous absorption. In this way the primary process is an absorption whose upper state is a dissociation-continuum.

An explanation of the results of Faltings, Groth and Harteck without any contradiction is only possible (Schmid and Gerö, 1937) on the basis of the dissociation scheme:

$$\mathrm{CO}(\mathrm{X}\,^1\Sigma) \rightarrow \mathrm{C}\,(^5S) + \mathrm{O}(^3P) - 89620\ \mathrm{cm.}^{-1}.$$

This dissociation scheme is founded on some predissociation effects, one of which has a limiting curve (Gerö, 1926) at 77497 ± 44 cm.$^{-1}$. Figure 2 contains this limiting curve. The predissociation which takes place along this curve is caused by molecular states belonging to the $\mathrm{C}(^1S)+\mathrm{O}(^3P)$ atomic term combination. The energy of the 1295 A. xenon line, 77185·9 cm.$^{-1}$, is not sufficient to raise a

CO molecule from its $X\,^1\Sigma(v=0,\ J=0)$ ground level over the limiting curve; but a considerable part of the molecule has at room temperature a rotational energy of some hundred cm.$^{-1}$. In figure 2 the rotational term sequence of the CO $X\,^1\Sigma$ $(v=0)$ state is drawn against $J(J+1)$ as abscissa. At the height of 77185·9 cm.$^{-1}$, a parallel with this term sequence is drawn, which intersects the limiting curve in the neighbourhood of $J=13$. Hence a CO molecule in the $X\,^1\Sigma$ $(v=0)$ ground state with $J \geqslant 13$ has quite enough rotational energy to dissociate on absorbing the 1295 A. xenon line. About 20% of all CO molecules are in states with $J \geqslant 13$ at room temperature; these molecules, absorbing the xenon radiation, rise in the dissociation continuum above the limiting curve, and as the mean life-time of this state is less than 10^{-13} sec., i.e., roughly one hundredth of the time interval between two collisions, the unit quantum efficiency of the dissociation process is quite well explicable.

From the $C(^1S)+O(^3P)$ atomic term combination, only triplet molecular terms arise. Transitions between the singlet ground state of the CO molecule and these triplet states are intercombination transitions. In the band spectrum of CO some intercombination band systems are known; one of them, the $A\,^3\Pi - X\,^1\Sigma$ (Cameron) band system, was photographed with fairly high intensity both in emission (Gerö, 1938) and absorption (Gerö, 1937). A number of perturbation and predissociation phenomena, caused by triplet states, were also observed in singlet states. For example, the predissociation in the $A\,^1\Pi$ molecular state at 77497 cm.$^{-1}$ is also caused by a triplet state; the intensity drop at this energy height is not very large and extends only over a few rotational levels of the $v=9$, 8 and 7 vibrational terms. Again, bands with $v'=10$ have rather high intensity. It follows that only above and close to the limiting curve is the probability of a transition without radiation sufficiently great to cause observable intensity diminutions. This is also the reason why no continuous absorption occurs at this energy level in the absorption spectrum of CO (see e.g. the absorption plates of Hopfield and Birge quoted in Bonhoeffer and Harteck, 1933). The effective region above the limiting curve is in fact too narrow, and the transition probability too small, to cause an observable absorption on plates taken with rather low resolving power.

Although the energy height of the ineffective 1470 A. xenon line—8·38 ev.—is about 1·5 ev. larger than the dissociation energy of CO, there is no atomic term-combination in the immediate neighbourhood of this energy height. Therefore, the CO molecule cannot absorb this wave-length. In this way no lower limit of a dissociation energy can be established by the non-appearance of a photochemical effect; a photodissociation possible on energy grounds only takes place if there is in the energy level of the absorbed radiation a continuum in which the transition probability from the ground level is large enough to give rise to considerable absorption.

REFERENCES

BACHER, R. F. and GOUDSMIT, S., 1932. *Atomic Energy States* (New York).

BONHOEFFER, K. F. and HARTECK, P., 1933. *Grundlagen der Photochemie*, p. 132.

FALTINGS, K., GROTH, W. and HARTECK, P., 1938. *Z. phys. Chem.* (B), **41**, 15.

GAYDON, A. G. and PENNEY, W. G., 1942. *Nature, Lond.*, **150**, 406.

GERÖ, L., 1936. *Z. Phys.*, **100**, 374; 1938. *Ibid.*, **109**, 204.

GERÖ, L., HERZBERG, G. and SCHMID, R. F., 1937. *Phys. Rev.*, **52**, 467.

HARTECK, P. and OPPENHEIMER, F., 1932. *Z. phys. Chem.* (B), **16**, 77.

SCHMID, R. and GERÖ, L., 1936. *Z. Phys.*, **101**, 343; 1937. *Z. phys. Chem.* (B), **36**, 105.

DISCUSSION

Dr. A. G. Gaydon. The comments by Schmid and Gerö on Gaydon and Penney's interpretation of the photodissociation of CO read plausibly, but are at variance with the direct experimental evidence on the absorption spectrum. This has been examined by Leifson (1926), using 15 mm. of gas at 1 atm. (i.e. half the path length used by Faltings, Groth and Harteck). His description and published spectrogram show clearly that there is no general absorption of appreciable strength in this region ; since the (10,0) band is about the same strength as the (9,0) and (11,0) bands [*] it is most unlikely that there is a localized region of continuous absorption around 1295 A. Schmid and Gerö mention a *half-value* width of, say, 0·1 cm.$^{-1}$ for the band lines. This may be correct, although pressure effects might increase it somewhat at 1 atm., but it is the *overall* width, i.e. the width over which absorption is appreciable, with which we are concerned. For absorption in the main resonance system by 30 mm. of gas at 1 atm. this must be very much greater. Indeed since the lines of the bands are spaced several cm.$^{-1}$ apart it would not have been possible for Leifson, who used low resolving power, to record the strong absorption which he in fact observed if the overall line width had not been considerable. The experimental evidence clearly proves that the only strong absorption by CO between 1250 and 1400 A. is due to the Fourth Positive band system, and any interpretation of the photodissociation must start from this established fact.

It is now generally realized that the low value for the dissociation energy of CO favoured by Schmid and Gerö cannot be reconciled with equilibrium measurements of the heat of sublimation of carbon, which indicate a very much higher value than the 74 k.cal/mole required by Schmid and Gerö's theory. This has recently been discussed at some length (Gaydon, 1946).

REFERENCES

Gaydon, A. G., 1946. *Dissociation Energies and Spectra of Diatomic Molecules* (London : Chapman and Hall, in the press).

Leifson, S. W., 1926. *Astrophys. J.*, **63**, 73.

---

[*] In Leifson's spectrogram there is some N_2 absorption superposed on the (8,0) and (9,0) bands of CO.

NOTE ON THE CALCULATION OF OPTICAL CONSTANTS

By D. J. PRICE,
Walthamstow

MS. communicated by H. Lowery, 13 *March* 1946

The calculation of the optical constants of an absorbing medium from the polarization produced in a beam of light reflected from the surface presents certain difficulties due to the somewhat cumbersome equations usually employed.

A new equation more suitable for routine computing is here developed and discussed with reference to previous work.

Refraction in an absorbing medium may be characterized by a complex refractive index, $N=n+ik$. This results in a complex angle of refraction χ, corresponding to a real angle of refraction ϕ, for by Snell's law

$$\sin\phi = N\sin\chi.$$

Further, if the incident light consists of two equal components in phase, and polarized in directions parallel and perpendicular to the plane of incidence, then after reflection the phase difference Δ and the ratio of the two amplitudes ρ will be given by Fresnel's equations as

$$-\rho e^{i\Delta}=\frac{\cos(\phi+\chi)}{\cos(\phi-\chi)}.$$

It is usual to measure Δ by restoring the phase difference by means of some type of compensator and then determining the azimuth of the resulting plane polarized light with respect to the plane of incidence. If this angle of azimuth be designated ψ, then

$$\rho=\tan\psi.$$

Combining the three equations, we have, after a simple reduction,

$$N^2=\left(\frac{1-Z}{1+Z}\right)^2\sin^2\phi\tan^2\phi+\sin^2\phi,$$

where

$$\begin{aligned}Z&=e^{i\Delta}\tan\psi\\&=(\cos\Delta+i\sin\Delta)\tan\psi.\end{aligned}$$

It is from this that most workers have proceeded. Drude (1889 to 1898), to whom the first analysis is due, developed two alternative methods.

First, for an approximate treatment, the $\sin^2\phi$ term may be neglected. This enables the root to be taken on both sides and therefore yields a simple expression for N. There are two objections to this approximation for modern work: (*a*) it is more inaccurate than is allowable with more recent experimental techniques, and (*b*) the theory of dispersion of metals involves N^2 and not N itself.

Secondly, Drude introduces three new, intermediate quantities defined by the equations :—

$$\begin{aligned}\tan Q&=\sin\Delta\tan 2\psi,\\\cos P&=\cos\Delta\sin 2\psi,\\S&=\sin\phi\tan\phi\tan\tfrac{1}{2}P.\end{aligned}$$

Taking $N^2=A+iB$, he obtains, after some reduction,

$$\begin{aligned}A&=S^2\cos 2Q+\sin^2\phi,\\B&=S^2\sin 2Q.\end{aligned}$$

This two-stage formula, although leading to a fairly simple form, entails quite a lengthy computation.

A simple and *direct* formula may be developed as follows :—

$$\begin{aligned}N^2&=\left(\frac{1-Z}{1+Z}\right)^2\sin^2\phi\tan^2\phi+\sin^2\phi\\&=\frac{2\sin^2\phi\tan^2\phi}{\frac{1}{2}\left(Z+\frac{1}{Z}\right)+1}+\tan^2\phi.\end{aligned}$$

But

$$\tfrac{1}{2}\left(Z+\frac{1}{Z}\right)+1=\frac{\cos\Delta}{\sin 2\psi}+\frac{\sin\Delta}{\tan 2\psi}.$$

Hence, if we put

$$X=\frac{\cos\Delta}{\sin 2\psi}+1$$

and

$$Y=\frac{\sin\Delta}{\tan 2\psi},$$

we derive

$$A=\tan^2\phi-2\sin^2\phi\tan^2\phi\left(\frac{X}{X^2+Y^2}\right),$$

$$B=2\sin^2\phi\tan^2\phi\left(\frac{Y}{X^2+Y^2}\right).$$

Further,

$$X^2+Y^2=(\operatorname{cosec}2\psi+\cos\Delta)^2.$$

Therefore,

$$A=\tan^2\phi-\frac{2\sin^2\phi\tan^2\phi\ (\cos\Delta+\sin2\psi)}{\sin2\psi(\cos\Delta+\operatorname{cosec}2\psi)^2},$$

$$B=\frac{2\sin^2\phi\tan^2\phi\sin\Delta}{\tan2\psi(\cos\Delta+\operatorname{cosec}2\psi)^2}.$$

This form is well suited to a series of computations based on observations made at different wave-lengths with a constant angle ϕ. In such a case the calculations may be set out in the following standard form :—

$\phi=73^\circ 0'$,
$\Delta=68^\circ 2\tfrac{1}{2}'$,
$2\psi=59^\circ 38'$,

$D=\tan^2\phi=10{\cdot}698$,
$C=\log(2\sin^2\phi\tan^2\phi)$
$=1{\cdot}29155$.

			logs	differences
		C	= 1·29155	0·92051 ......(*a*)
			0·37104 ↑ ×2	
cosec 2ψ	=	1·15900		
sum		1·53292 →	0·18552	
cos Δ	=	0·37392		
sum		1·23673 →	0·09228	0·15636 ..... (*b*)
sin 2ψ	=	0·86281 →	$\bar{1}$·93592	
		log sin Δ	= $\bar{1}$·96729	$\bar{1}$·73512 ......(*c*)
		log tan 2ψ	= 0·23217	
			antilogs	
a+*c*		0·65563 →	4·525	= *B* = 2*nk*
a+*b*		1·07687 →	11·936	
			10·698	= *D*
		diff.	1·238	= −*A* = k^2-n^2

It was found in practice that the substitution of this method of calculation for the previously noted formula of Drude involved a reduction of computing time by a factor of about two.

REFERENCES

Drude, 1889. *Ann. Phys., Lpz.*, **36,** 532 ; 1890. *Ibid.*, **39,** 481 ; 1898. *Ibid.*, **64,** 159.

ULTRA-VIOLET ABSORPTION BAND-SYSTEMS OF SnSe AND SnTe

BY E. E. VAGO AND R. F. BARROW,
Physical Chemical Laboratory, University of Oxford

MS. received 1 *July* 1946

ABSTRACT. The study of the absorption spectra of SnSe and of SnTe (Walker, Straley and Smith, *Phys. Rev.* **53,** 140 (1938); Barrow and Vago, *Proc. Phys. Soc.* **56,** 78 (1944)) has been extended to cover the ultra-violet region down to 2100 A. New systems of both molecules have been analysed vibrationally. The results, including those derived from the D–X systems already known, are summarized in the following table:

Molecule	State	ν_e	ω_e	$x_e\omega_e$	$y_e\omega_e$	D (ev.)
SnSe	F	(47850)	(290)	—	—	—
	E	30738·9	196·6	−0·77	−0·0016	0·7
	D	27549·6	225·1	−0·69	—	1·1
	X	0	331·2	$-0{\cdot}73_5$	—	4·5
SnTe	G	(47260)	(230)	—	—	—
	F	44033·5	229·7	−1·25	−0·003	—
	E	(28000)	(150)	—	—	0·6
	D	25444·3	179·1	−0·40	—	0·9
	X	0	259·5	−0·50	—	4·1

§ 1. INTRODUCTION

THE results of some work on the emission and absorption spectra of SnSe and SnTe were reported in a discussion on band-spectra (Barrow, 1944). However, the low-wave-length limit of these absorption studies was about 3700 A., and there were several reasons which led us to suspect the existence of at least one further absorption system (E–X) of both molecules at shorter wave-lengths. For, not only were there well-known E–X systems of SnO and SnS, but also bands which apparently belonged to these systems had already been identified in the emission spectra of SnSe and SnTe (Barrow, 1940; Barrow and Vago, 1943).

We decided therefore to examine the absorption of SnSe and SnTe vapours in the region 2100–3800 A. (at temperatures of the order of 1000° C.). This proved to be more fruitful than we had expected.* We have been able not only to confirm the existence of the E–X systems of these molecules (Barrow and Vago, 1944), but also, for SnSe, to carry out the vibrational analysis of the rather extensive band-system with v'-assignments going from 0 to 29: this leads to a fairly reliable estimate of the dissociation energy for the E state of SnSe. Unfortunately the absorption spectrum of SnTe in the corresponding region is rather complicated, and we have failed to produce a convincing vibrational scheme: for the E–X system of this molecule we are certain, therefore, only of the approximate position of the E level relative to the ground state.

---

* While this work was being prepared for publication a note on the same subject was published by D. Sharma (*Nature*, **157,** 663 (1946)). His results for SnSe and SnTe appear to be similar to ours, but his interpretation differs in detail. Further comment may be postponed until a full account of his work is available.

At shorter wave-lengths than the E–X systems we have found two new systems of SnTe and one of SnSe. Of these, the F←X system of SnTe was quite well developed, and the vibrational analysis presented few difficulties. We have obtained only fragmentary information about the G←X system of SnTe and the F←X system of SnSe, as the low-λ portions of these systems lie beyond the limit of the present measurements. Corresponding systems of SnO and SnS may be expected in the vacuum ultra-violet region (1600–2000 A.), but no molecule of the group—apart from the first member, CO—has yet been studied in this region: there are several indications that such work would be worth while.

The results obtained with SnSe and SnTe suggested an examination of the absorption spectra of the analogous lead compounds, PbO, PbS, PbSe and PbTe, in the same region. This we have done, and have found new systems of the last three compounds: the results will be reported in a later paper. The experimental methods have been described in earlier papers in this series; we therefore proceed to consider the results.

§ 2. RESULTS FOR SnSe

Ultra-violet absorption bands attributed to SnSe have been found in two regions, 4150–2730 A. and 2200–<2070 A. These bands have been assigned to three systems which are discussed below. Plates taken at somewhat lower pressures of SnSe than are required for the best development of the bands in the region 4150–2730 A. show what appears to be continuous absorption at $\lambda < 2730$ A.

(i) The D–X system: λλ3410–4150 A.

A vibrational analysis of this sytem, based on emission spectrograms, was given some time ago (Barrow and Vago, 1943). We have now confirmed this analysis by measurements of absorption plates taken in a first order of the 2·4-m. concave-grating instrument (linear dispersion 7·4 A./mm.). The positions of a number of new bands assigned to this system have been located. Both sets of measurements are summarized in a Deslandres scheme (table 1): improved values of the band-head constants derived from table 1 are given in table 7.

(ii) The E–X system: λλ2730–3410 A.

A spectrogram of the bands assigned to this system, taken with a large quartz Littrow spectrograph (dispersion ~5 A./mm. at 3250 A.), is reproduced in figure 1. Significant are (1) the long, rather regular, progression of bands in the region 2850–3050 A., (2) the sharpness of the band-heads around 3250 A., and (3) the overlap with bands of the D–X system at about 3400 A. Of these features, (1) gives immediately the order of magnitude of ω_e', since these bands are most likely to constitute a progression with $v''=0$, and (2) provides an approximate value for ν_e, since the diffuseness of the outlying band-heads may be attributed to the vibrational isotope effect (Barrow and Vago, 1943). Since, in addition, the ground-state vibrational terms for SnSe are already known quite accurately, it will be clear that there cannot be many different ways of assigning plausible v', v'' values to the bands. We have found only one such way (table 2). This analysis leads to the values of the constants listed in table 7. It should be

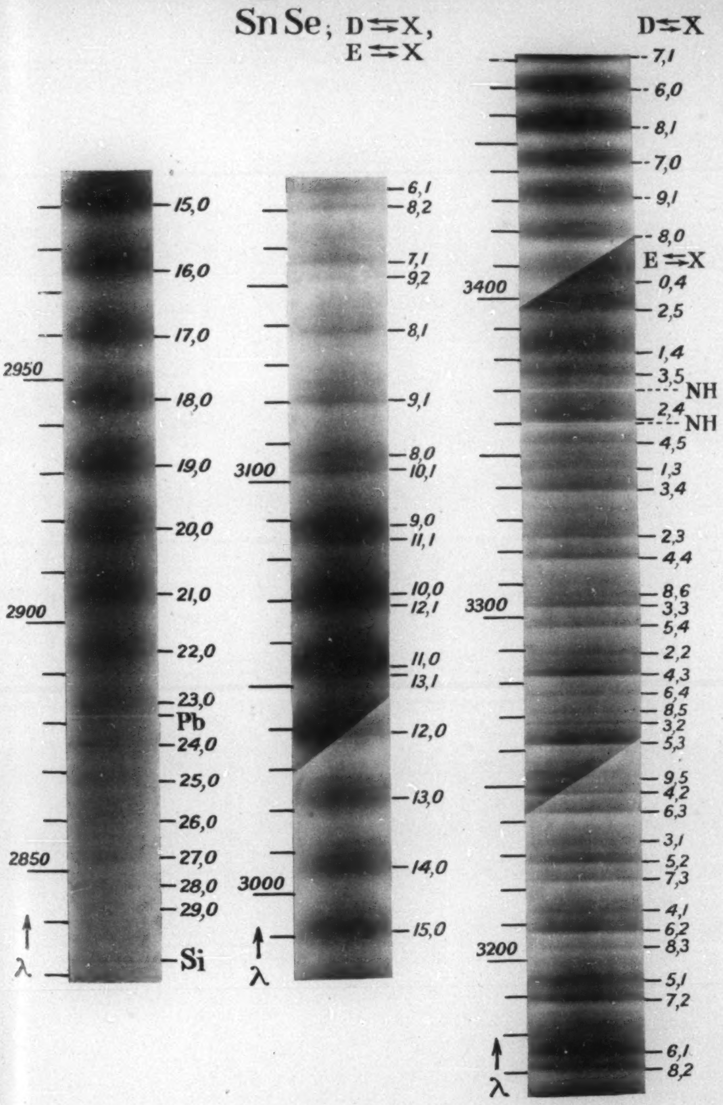


Figure 1. Bands of the D–X and E–X systems of SnSe in absorption (quartz Littrow spectrograph).

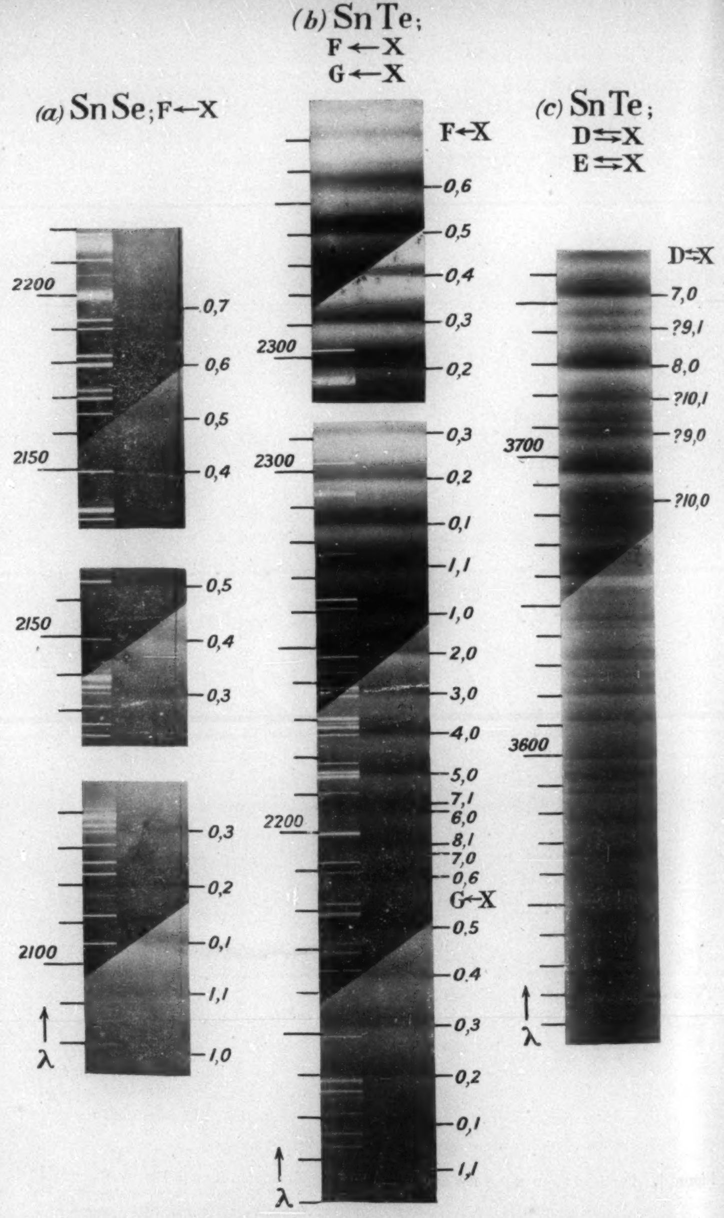


Figure 2. (*a*) E←X system of SnSe.
(*b*) F←X and G←X systems of SnTe.
} Hilger medium quartz spectrograph.
(*c*) D–X and E–X systems of SnTe. 2·4-m. grating spectrograph.
The reference lines in (*a*) and (*b*) are Cu arc lines.

Table 1. Band-head data for the D⇌X system of SnSe

Wave-numbers, cm.$^{-1}$; **bold-face** denotes bands measured both in absorption and in emission ;
large Roman, bands measured in absorption only;
large Roman*, bands measured in emission only.

↑ v' \ v'' →	0		1		2		3		4		5		6		7		8		9		10		11		12
8	29249·4* *3417·90*	330·8	28918·6* *3457·00*																						
	207·9		205·9																						
7	29041·5* *3442·37*	328·8	28712·7* *3481·78*																						
	221·5		218·3																						
6	28820·0* *3468·82*	325·6	28494·4* *3508·46*								27185·6 *3677·37*														
	219·0		224·2																						
5	28601·0* *3495·38*	330·8	28270·2* *3536·29*																						
	217·6		215·8																						
4	28383·4* *3522·18*	329·0	28054·4* *3563·49*	328·3	**27726·1** *3605·69*	326·9	27399·2 *3648·70*	325·7	27073·5 *3692·60*						26106·1 *3829·44*										
	218·4		217·8		218·9		220·4		220·0																
3	28165·0* *3549·50*	328·4	**27836·6** *3591·37*	329·4	**27507·2** *3634·38*	328·4	**27178·8** *3678·30*	325·3	26853·5 *3722·86*	324·7	26528·8 *3768·42*														
	220·4		221·4		222·0		221·1		221·9		219·3														
2	**27944·6** *3577·49*	329·4	**27615·2** *3620·17*	330·0	**27285·2** *3663·95*	327·5	**26957·7** *3708·47*	326·1	**26631·6** *3753·88*	322·1	26309·5 *3799·83*	321·4	25988·1 *3846·83*								24709·2 *4045·94*	311·5	24397·7 *4097·60*	314·5	24083·2 *4151·11*
			223·6		222·5		221·6		220·7		222·7										218·8		223·4		
1	—		**27391·6** *3649·72*	328·9	**27062·7** *3694·18*	326·6	**26736·1** *3739·20*	325·2	**26410·9** *3785·24*	324·1	26086·8 *3832·28*				25444·1 *3929·07*	319·2	**25124·9** *3979·00*	318·7	**24806·2** *4030·12*	315·8	24490·4 *4082·08*	316·1	24174·3 *4135·46*		
			225·0		223·7		224·6		225·2		225·0				222·9		222·9		222·5						
0	—		27166·6 *3679·95*	327·6	**26839·0** *3724·87*	327·5	**26511·5** *3770·88*	325·8	**26185·7** *3817·80*	323·9	**25861·8** *3865·62*	321·7	**25540·1** *3914·31*	319·1	**25221·0** *3963·83*	319·0	**24902·0** *4014·61*	318·3	24583·7 *4066·59*						

Table 2. Band-head data for the E–X system of SnSe

e indicates band also observed in emission.

* band possibly present, but overlaid by the 7,0 band of the D–X system.

v′	v″ → 0		1		2		3		4		5		6
29	35175 *2842·1*												
	57												
28	35118 *2846·7*												
	74												
27	35044 *2852·7*												
	88												
26	34956 *2859·9*												
	90												
25	34866 *2867·3*												
	92												
24	34774 *2874·9*												
	103												
23	34671 *2883·4*												
	126												
22	34545 *2893·9*												
	140												
21	34405 *2905·7*												
	149												
20	34256·4 *2918·31*												
	155·3												
19	34101·1 *2931·60*												
	160·9												
18	33940·2 *2945·50*												
	164·5												
17	33775·7 *2959·84*												
	170·3												
16	33605·4 *2974·85*												
	167·4												
15	33438·0 *2989·74*												
	171·5												
14	33266·5 *3005·15*	327	32930 *3035·9*										
	175·8		181										
13	33080·7 *3022·03*	331·6	32749·1 *3052·61*										
	174·4		173·1										
12	32906·3 *3038·05*	330·3	32576·0 *3068·86*										
	177·8		177·6										
11	32728·5 *3054·56*	330·1	32398·4 *3085·68*										
	178·0		177·0										
10	32550·5 *3071·26*	329·1	32221·4 *e* *3102·63*										
	181·2		179·4										
9	32369·3 *3088·45*	327·3	32042·0 *e* *3119·99*				31387·5 *3185·06*				30745·3 *3251·60*		
	180·9		181·9				181·0				182·9		
8	32188·4 *3105·81*	328·3	31860·1 *3137·82*	328·5	31531·6 *3170·51*	325·1	31206·5 *e* *3203·54*				30562·4 *3271·06*	322·1	30240·3 *3305·90*
			183·9		182·8		183·7						
7			31676·2 *3156·03*	327·4	31348·8 *e* *3189·00*	326·0	31022·8 *3222·51*	325·8	30697·0 *3256·71*				
			187·4		188·3		188·0		186·4				
6			31488·8 *3174·83*	328·3	31160·5 *3208·26*	325·7	30834·8 *e* *3242·15*	324·2	30510·6 *3276·61*				
			190·0		189·5		189·4		190·5				
5			31298·8 *3194·51*	327·8	30971·0 *e* *3227·90*	325·6	30645·4 *3262·18*	325·3	30320·1 *e* *3297·19*				
			189·1		188·9		189·2		191·5				
4	31441·1 *3179·64*	331·4	31109·7 *e* *3213·51*	327·6	30782·1 *e* *3247·69*	325·9	30456·2 *e* *3282·42*	327·6	30128·6 *e* *3318·16*	322·2	29806·4 *3354·02*		
	189·5		187·9		190·6		191·4		187·9		188·8		
3	31251·6 *3198·89*	329·8	30921·8 *3233·03*	330·3	30591·5 *e* *3267·95*	326·7	30264·8 *3303·22*	324·1	29940·7 *e* *3338·98*	323·1	29617·6 *e* *3375·40*		
			189·5		190·9		191·0		192·8		191·9		
2			30732·3 *3252·97*	331·7	30400·6 *3288·46*	326·8	30073·8 *3324·00*	325·9	29747·9 *e* *3360·62*	322·2	29425·7 *e* *3397·42*	320·8	29104·9 *e* *3434·87*
			194·9		192·5		192·0		190·2		196·3		
1			30537·4 *3273·72*	329·3	30208·1 *3309·42*	326·3	29881·8 *3345·55*	324·1	29557·7 *3382·25*	328·3	29229·4 *3420·25*		
							197·6		197·7				
0							29684·2 *3367·83*	324·2	29360·0 *3405·44*		*		

noted, however, that the expression for G_v' derivable from table 7 is valid only up to $v'=17$: for the representation of higher-v' levels, additional terms in higher powers than $(v+\frac{1}{2})^3$ would be necessary. The biggest uncertainty in the analysis attaches to the reality of the bands with $v'=0$. The difficulties here are exactly analogous to those encountered in the analysis of the E–X system of SiS, in which there is a rather similar intensity distribution (Vago and Barrow, 1946). For SnSe, however, there is the additional confirmation from the vibrational isotope effect that the scheme given in table 2 is correct.

(iii) The F←X system: $\lambda\lambda < 2070$–2200.

Only a part of this system has been observed (figure 2, table 3). Most of the bands appear to form a progression, probably with $v'=0$. That this is a reasonable conclusion may be seen by comparing the intensity distribution with that for the F–X system of SnTe (figure 2). The wave-number differences then identify the absorbing molecule as SnSe, and lead to a provisional assignment of v'' values which can hardly be in error by more than ± 2 units and may well be correct. The estimate of the upper state frequency given in table 7 rests on measurements of only two weak bands: it accords satisfactorily with what might be expected by analogy with the F-state data for SnTe, but the figure given cannot be regarded as much more than a reasonable guess.

Table 3. Band-head data for the F←X system of SnSe

Wave-length (I.A.)	ν, cm.$^{-1}$	$\Delta\nu$	Provisional v', v'' values
2077·5	48119		1,0
		333	
2092·0	47786		1,1
		288	
2104·7	47498		0,1
		328	
2119·3	47170		0,2
		324	
2134·0	46846		0,3
		325	
2148·9	46521		0,4
		323	
2163·9	46198		0,5
		326	
2179·3	45872		0,6
		324	
2194·8	45548		0,7
2080·5	48050		?
2096·6	47681		?

§ 3. RESULTS FOR SnTe

The banded ultra-violet absorption spectrum of SnTe can also be divided into two regions: 3075–4200 A. and <2120–2360 A. There is some evidence of a region of continuous absorption at $\lambda < 3075$ A.

(i) The D–X system: λλ 3700–4200 A.

The vibrational structure of this system has been adequately discussed in earlier papers. Since, however, there is a region where the long-λ end of the E–X system overlaps the D–X system, we did make a particular attempt to sort out the bands which might be assigned to high v' values in the D–X system (3680–3750 A.: figure 2, table 4). We were not able to draw any definite conclusions.

Table 4. Band-heads of the E⇌X system of SnTe

Wave-length (I.A.)	Wave-number, cm.$^{-1}$ (*in vac.*)	Intensity	Wave-length (I.A.)	Wave-number, cm.$^{-1}$ (*in vac.*)	Intensity	
3327·2	30046	–	*3540·82*	28234·0*	1	
35·5	29972	–	*44·51*	204·6*	2	
42·1	913	–	*54·54*	125·0*	2	
50·4	840*	–	*59·13*	088·8*	1	
65·5	705*	–	*65·63*	037·6*	3	
74·1	629*	–	*76·66*	27951·1*	4 *e*	
80·3	575*	–	*78·35*	937·5*	1	
88·5	503	–	*87·17*	869·2*	5 *e*	
95·6	442*	–	*91·88*	832·7*	2 *e*	
3403·8	371*	–	*98·34*	782·7*	7 *e*	
11·7	303	–	*3611·31*	682·9*	6 *e*	
19·7	234*	–	*20·28*	614·3*	3	
27·5	167	–	*26·04*	570·5*	2 *e*	
34·5	108*	–	*31·73*	527·3*	5	
43·9	029*	–	*42·15*	448·5	2 *e*	
50·5	28973*	–	*45·07*	426·5*	4 *e*	
58·8	903*	1	*59·82*	316·0*	9 *e*	
68·0	827	1	*63·49*	288·6	1	
74·5	773*	0	*73·14*	216·9	2	
83·1	702*	2	*76·36*	193·1*	2 *e*	
91·5	633*	0	*84·64*	132·0*	1	? D⇌X, 10,0
99·46	567·7*	2	*94·04*	063·0*	10 *e*	
3509·08	489·4*	3	*3706·53*	26971·8*	3	? D⇌X, 9,0
12·1	465*	0	*11·17*	938·0*	3 *e*	
14·6	444*	0	*18·38*	885·8*	4	? D⇌X, 10,1
16·9	426	0	*29·03*	809·0*	10 *e*	D⇌X, 8,0
22·94	377·3*	1	*41·74*	718·0*	3	? D⇌X, 9,1
27·01	344·6*	0	*46·65*	682·9*	3 *e*	
31·05	312·1*	1				
34·79	282·2	2				

e indicates a band also observed in emission.
* indicates a band which may be fitted into a v' progression.

(ii) The E–X system: λλ 3075–3720 A.

When bands of this system were first observed—in emission—an attempt was made to carry out a vibrational analysis (Barrow, 1940). That only a tentative result could be given was attributed to the incompleteness of development of the system under these conditions. The new absorption plates (figure 2) are much superior to the emission ones, and we are satisfied that the band-head measurements given in table 4 are reliable. Nevertheless, we have not been able to assign the bands to any really plausible vibrational scheme. One difficulty—that the long-λ end of this system overlaps the short-λ end of the D–X system—has been mentioned. Others are that the value of $x_e''\omega_e''$ for SnTe (0·5 cm.$^{-1}$) is so small as to preclude the unambiguous assignment of v'' values from wave-number differences alone, and that neither sequences nor progressions are particularly well marked. Over a considerable range—some 1000 to 2000 cm.$^{-1}$ to the short-λ side of the origin—a frequency difference of about 135 is common, suggesting that ω_e' may be about 150 cm.$^{-1}$. We have, however, only succeeded in assigning the bands to a scheme involving this frequency by assuming considerable and rather numerous upper-state perturbations, about which we have no independent information.

The vibrational isotope effect enables the system-origin to be fixed at 28000 ± 250 cm.$^{-1}$ (3570 A.), and although bands can be followed down to about 3075 A., we have no evidence that there is more than one upper electronic level of SnTe in the energy range corresponding to these wave-lengths.

(iii) The F←X system: λλ 2190–2360.

This system is bounded by two strong progressions (figure 2), and there can be little doubt about the correctness of the analysis given in table 5.

Table 5. Band-head data for the F←X system of SnTe

v' ↑ / v'' →	0		1		2		3		4		5		6
8	45759 *2184·7*	259	45500 *2197·1*										
	205		204										
7	45554 *2194·5*	258	45296 *2207·0*				44779·3 *2232·48*						
	210						208·6						
6	45343·9 *2204·68*						44570·7 *2242·93*						
	213·9												
5	45130·0 *2215·13*				44611·4 *2240·88*				44100·0 *2266·87*				
	217·6				217·4								
4	44912·4 *2225·86*				44393·0 *2251·91*	255·5	44137·4 *2264·95*						
	221·0						216·4						
3	44691·4 *2236·87*						43921·0 *2276·11*						
	219·3												
2	44472·1 *2247·90*	259·6	44215·5 *2261·10*										
	224·9		224·1										
1	44247·2 *2259·33*	258·8	43988·4 *2272·62*	259·1	43729·3 *2286·09*								
			229·4		227·9								
0	—		43759·0 *2284·54*	257·6	43501·4 *2298·07*	256·1	43245·3 *2311·68*	255	42990 *2325·4*	258	42732 *2339·4*	253	42479 *2353·4*

(iv) The G←X system: $\lambda\lambda < 2120$–2190.

Both relative intensities and the vibrational isotope effect show that the bands in the region 2120–2360 A. belong to two different systems (figure 2). This is confirmed by the vibrational analysis. Most of the observed bands of the G←X system fit into what appears to be a $v' = 0$ progression. The wave-number differences identify the absorbing molecule as SnTe. Possible values of ν_e and ω_e for the G level are given in table 7: they are subject to the same kinds of uncertainty as were discussed in connection with the F←X system of SnSe.

Table 6. Band-head data for the G←X system of SnTe

Wave-length (I.A.)	ν, cm.$^{-1}$	$\Delta\nu$	Provisional v', v'' values
2117·5	47211		1,1
		227	
2127·7	46984		0,1
		257	
2139·4	46727		0,2
		256	
2151·2	46471		0,3
		260	
2163·3	46211		0,4
		259	
2175·5	45952		0,5
		258	
2187·8	45694		0,6

Table 7. Summary of band-head constants

Molecule	State	ν_e	ω_e	$x_e\omega_e$	$y_e\omega_e$
SnSe	F	(47850)	(290)	—	—
	E	30738·9	196·6	−0·77	−0·0016
	D	27549·6	225·1	−0·69	—
	X	0	331·2	−0·736	—
SnTe	G	(47260)	(230)	—	—
	F	44033·5	229·7	−1·25	−0·003
	E	(28000)	(150)	—	—
	D	25444·3	179·1	−0·40	—
	X	0	259·5	−0·50	—

§ 4. DISCUSSION

It is interesting to compare the new results for SnSe and SnTe with those for SnO and SnS. The relevant values of ν_e, and of the force constant expressed as percentages of those for the ground states, are given in table 8. The ν_e values, both for the D and for the E states, decrease fairly regularly from oxide to telluride—as might have been expected. They suggest that the states labelled with a common letter may be spectroscopically similar, and this suggestion receives some support from the values of the force constants. Thus the D states have force constants which are all about 48% of those for the ground states, while the figures for the E levels are all close to 35%. This behaviour is quite analogous to that displayed by the D and E states of the corresponding silicon compounds (Vago and Barrow, 1946).

Table 8. Values of ν_e, and of the force constants expressed as percentages of those for the ground states

Molecule : State	SnO	SnS	SnSe	SnTe
G	—	—	—	(47260) (78%)
F	—	—	(47850) (77%)	44034 78·4%
E	36803 36·6%	33037 36·4%	30739 35·1%	(28000) (33%)
D	29625 50·2%	28338 46·3%	27550 46·2%	25444 47·6%

The figures in table 8 also suggest that F and G levels of SnO and SnS and a G level of SnSe should exist. The force constants for these states may be about 78% of those in the ground states, so that $\omega_e'(\text{SnO}) \sim 730$, $\omega_e'(\text{SnS}) \sim 430$ cm.$^{-1}$. It is hardly possible to make any accurate prediction of the positions of the absorption systems involving these states, but they should lie in the vacuum ultra-violet region within the rough limits 1600 to 2000 A.

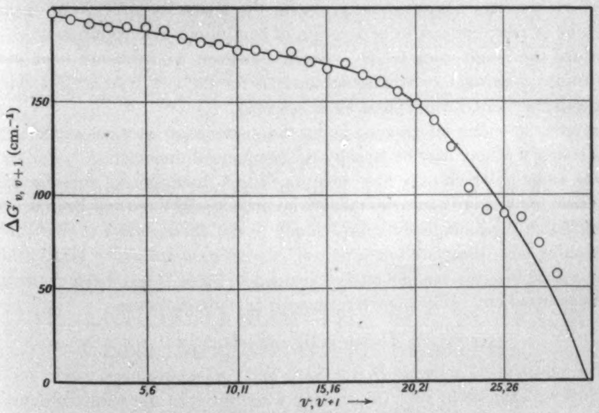


Figure 3. The variation of vibrational interval with vibrational quantum number for the E state of SnSe.

It remains to consider the dissociation energies for these states. The most reliable figure is that for the E level of SnSe, where bands up to $v' = 29$ have been observed (table 2): in this region the vibrational interval between successive levels is less than 30% of ω_e', so that a fairly good estimate of D_E should be obtained by integration of the curve got by plotting $\Delta G'_{v,\,v+1}$ against v, $v+1$. This plot is shown in figure 3: from it D_E is found to be 4800 cm.$^{-1}$. It is noteworthy that this figure is only 40% of that obtained by linear extrapolation of the first fifteen vibrational intervals.

The integrated value of D_E might be in error either by faulty location of the band-heads or by incorrect assignment of the v' values. The latter point has been discussed above: it is difficult to estimate the error arising from the uncertainty in the band-head measurements, for which the vibrational isotope effect is mainly responsible. For SnSe this effect is so complicated that we can hardly hope for detailed resolution of the various heads of a given band, while at the same time its magnitude is such that the total spread of a band distant 1000 cm.$^{-1}$ from the origin is nearly 25 cm.$^{-1}$ In these circumstances all that can be done is to measure throughout the strongest bands. This we have attempted to do, and it seems likely that the resulting limits of error of the integrated value of D_E are not very large. However, another figure for D_E may be derived by measuring the point at which it appears that the banded absorption passes over into a continuum. It is difficult to estimate the position of this point accurately, but it seems to lie at about 2730 A., giving $D_E = 5900$ cm.$^{-1}$, or 0·73 ev. The fact that this figure is a little higher than that got by integration of the curve in figure 3 suggests that the vibrational levels do not in fact converge quite so rapidly at high v-values as has been drawn. Nevertheless, the agreement between the two values of D_E may be regarded as fairly satisfactory, and we can be reasonably confident that the true value is close to the higher figure, 0·73 ev., that is, about 50% of the value obtained by linear extrapolation.

No such certain figure can be given for the E state of SnTe, but again the beginning of what appears to be a region of continuous absorption may be used to deduce the rough value 0·6 ev. Since, however, we have not been able to put forward a satisfactory vibrational analysis for the E–X system, this estimate of D_E must be treated with considerable reserve.

In order to obtain information about the dissociation energies in the ground and D states, it is necessary to consider the products of dissociation. For various reasons, some of which have been mentioned in a discussion of the dissociation energies of the X, D and E states of the silicon compounds (Vago and Barrow, 1946), we think that the most likely interpretation of the present data is that the X, D and E states each dissociate into a pair of ground state atoms, $^3P + {}^3P$, although other possibilities are not definitely excluded. Table 9 has been constructed on this assumption. If it is correct, then for any one molecule

$$D_X = \nu \quad + D_D = \nu_E + D_E.$$

The values given above of D_E for SnSe and SnTe have been used: for SnO and SnS we have taken 50% (by analogy with SnSe) of the values obtained by linear extrapolation of the vibrational intervals. The values given for D_D are in all cases 70% of the linearly extrapolated values; these appear to us on general grounds to be reasonable, if somewhat arbitrary, estimates. For the ground states, for which the extrapolation is likely to be least in error, we have taken 100% of the linear values.

The quantities $\nu_E + D_E$, $\nu_D + D_D$ and D_X are given in rows v, vi, and vii of table 9. The agreement between them is fair—probably as good as can be expected. The least satisfactory value appears to be that of D_D for SnTe, which seems to be a good deal too high. The measure of concordance of the other values for a given molecule is some support for the view that the products of dissociation are the same in the three states, and leads us to attempt to derive improved values, first of D_X (for SnO and SnS the means of v, vi and vii have been taken; for SnTe the mean of v and vii; and for SnSe the value of D_E determined from the beginning of continuous absorption has been accepted), and thence of D_D and D_E. These improved values are given in rows viii, ix and x. As was found for the silicon compounds, there is a more or less regular decrease in the value of D for a given state on going from oxide to telluride, and of course in the values of D_X, D_D and D_E for a given molecule : the force constants follow similar general trends.

Table 9. Electronic energies and energies of dissociation (ev.)

		SnO	SnS	SnSe	SnTe
i	ν_E	4·56	4·10	3·81	$3{\cdot}4_7$
ii	ν_D	3·67	3·51	3·42	3·15
iii	D_E	1·50 *	1·17 *	0·73 †	(0·6) †
iv	D_D	2·39 ‡	1·91 ‡	1·59 ‡	1·74 ‡
v	ν_E+D_E	6·06	5·27	4·54	(4·07)
vi	ν_D+D_D	6·06	5·42	5·01	4·89
vii	D_X	5·62 §	5·50 §	4·62 §	4·17 §
Improved values					
viii	D_X	5·91	5·40	4·54	4·1
ix	D_D	2·24	1·89	1·12	0·9
x	D_E	1·35	1·30	0·73	0·6

* Obtained by taking 50% of the figures determined by linear extrapolation of the vibrational intervals.

† From an estimate of the onset of continuous absorption.

‡ obtained by taking 70% of the figures determined by linear extrapolation of the vibrational intervals.

§ Obtained by linear extrapolation of the vibrational intervals.

REFERENCES

Barrow, R. F., 1940. *Proc. Phys. Soc.*, **52**, 380.
Barrow, R. F., 1944. Discussion on Band Spectra. *Proc. Phys. Soc.*, **56**, 204.
Barrow, R. F., and Vago, E. E., 1943. *Proc. Phys. Soc.*, **55**, 326 ; 1944. *Ibid.*, **56**, 78.
Vago, E. E. and Barrow, R. F., 1946. *Proc. Phys. Soc.*, **58**, 538.

RADAR ECHOES FROM THE SEA SURFACE AT CENTIMETRE WAVE-LENGTHS

By H. DAVIES and G. G. MACFARLANE,
Telecommunications Research Establishment

Communicated by D. Taylor ; *MS. received* 31 *May* 1946

ABSTRACT. An account is given of quantitative measurements of the echo obtained from the sea surface with radars operating at wave-lengths of 1·25, 3 and 10 cm. under various weather conditions from calm to very rough. The necessary theoretical treatment is given, and a function $f(\theta)$, called the *scattering coefficient* of the surface, is defined.

§ 1. INTRODUCTION

The detection of small objects situated on the surface of the sea is limited by the fact that the sea in the vicinity of the target scatters backwards a certain amount of the incident energy, and the target echo is frequently masked by this scattered radiation. To effect any improvement in the discrimination of target from " clutter ", a knowledge of the magnitude of sea clutter for different wave-lengths and various states of the sea, and its dependence on angle of depression and range, is essential. An understanding of the scattering mechanism coupled with a theoretical treatment of the subject would be desirable.

A theoretical attack on the problems has been made,* in which it is assumed that the surface of the sea consists of cylindrical sinusoidal waves, the wave-length and wave-height of which both satisfy certain statistical distributions. The model of the sea is a two-dimensional one. However, the theoretical approach is handicapped by inadequate knowledge of the exact wave-form of the sea waves and the shape of the statistical distributions. A mathematical treatment based on any wave-shape other than the simple sinusoid is extremely difficult, and if a solution is obtainable it is well nigh incapable of interpretation.

In this paper emphasis is laid on the experimental results rather than on the theoretical problem.

§ 2. THEORETICAL CONSIDERATIONS

The notation given below is the one normally employed in radar, and will be adopted throughout this article

Let P = Peak power output of transmitter.

G = Power gain of aerial system over an isotropic radiator.

A = Effective aperture of aerial system $\left(G=\frac{4\pi A}{\lambda^2}\right)$.

λ = Wave-length of radar.

τ = Pulse length (in time).

c = Velocity of E.M. waves in air.

h = Height of radar above the sea.

R = Slant range.

θ = Angle of depression from horizontal. } See figure 1.

ϕ = Angle of azimuth from centre of beam. } See figure 1.

$F(\theta, \phi)$ = Amplitude polar diagram factor of aerial.

Φ = Beam width in azimuth, defined as the angular distance in radians between directions 3 db. down on the maximum on either side for one-way transmission.

Referring to figure 1, consider a small element of sea ds situated at P, at range R, depression θ, and azimuth ϕ.

The power density at P is given by

$$S_i=\frac{PGF^2(\theta, \phi)}{4\pi R^2}. \qquad \ldots\ldots(1)$$

The power intercepted by the element ds is $S_i \sin\theta \, ds$. If this power were scattered isotropically over the hemisphere above ds, then the power received by the radar would be

$$dP_r=\frac{S_i \sin\theta \, ds \, AF^2(\theta, \phi)}{2\pi R^2} \qquad \ldots\ldots(2)$$

$$=\frac{PGA \sin\theta \, F^4(\theta, \phi) \, ds}{8\pi^2 R^4}. \qquad \ldots\ldots(3)$$

---

* Unpublished paper by G. G. Macfarlane.

This formula has been derived on the assumption that the element *ds* behaves as a " perfectly rough " surface, i.e. a surface which scatters all the incident energy uniformly over the hemisphere above it. This definition of a " perfectly rough " surface is contrary to the optical concept of such a surface, which scatters according to Lambert's cosine law, but it is more convenient as a reference level or standard surface.

In general it is not true to say that *ds* scatters isotropically, and we must write

$$dP_r = \frac{PGA \sin\theta F^4(\theta, \phi) f(\theta)\, ds}{8\pi^2 R^4}, \qquad \ldots\ldots(4)$$

where $f(\theta)$ is also a function of the radar wave-length and the state of the sea prevailing. $f(\theta)$ may be called the scattering coefficient, and is essentially a ratio.

The expression (4) is the power received back from the element of sea *ds*. By the very nature of the pulse technique employed in radar, echoes are received by the radar at the same time from all elements in an area which are at ranges lying between R and $R+\tau\mathbf{c}/2$. This area is called the *resolvable area* of the radar. We assume that the surface behaves as an incoherent scatterer, and so dP_r must be summed or integrated over the resolvable area.

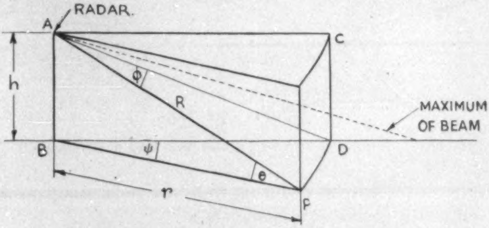


Figure 1. The configuration and the co-ordinates.
(*All figures Crown Copyright.*)

The resolvable area in this case for range R is the annulus between the two circles on the sea of slant radius R and $R+\tau\mathbf{c}/2$. Assuming a symmetrical beam, most of the energy is concentrated in a cone of semi-vertical angle $\Phi/2$. It can be seen that we have essentially two different configurations with which to deal:

1. The case in which the resolvable area is narrow compared with the area illuminated by the main part of the beam. See figure 2.

2. The case in which the resolvable area is wide compared with the area illuminated by the main part of the beam. See figure 3.

We shall deal separately with these two cases.

§ 2.1.

The mathematical expression for this state of affairs is

$$\frac{\tau\mathbf{c}}{2} \sec\theta \ll R\Phi \operatorname{cosec}\theta.$$

The maximum return from the resolvable area at range R will be given when this area lies in the centre of the main part of the beam as shown in figure 2. In this case we may proceed thus. As the element of sea, we may take

$$ds = \frac{\tau\mathbf{c}}{2}\sec\theta \,.\, r\,d\phi \qquad \ldots\ldots(5)$$

$$= \frac{\tau\mathbf{c}}{2} R\sec\theta\,d\phi.$$

$$\therefore\quad dP_r = \frac{PGA}{16\pi^2 R^4} R\tau\mathbf{c}\tan\theta \,.\, F^4(\theta,\phi) f(\theta)\,d\phi, \qquad \ldots\ldots(6)$$

$$P_r = \frac{PGA}{16\pi^2 R^3}\tau\mathbf{c}\tan\theta f(\theta)\int F^4(\theta,\phi)\,d\phi. \qquad \ldots\ldots(7)$$

The area of integration is small enough to justify the removal of $R\tan\theta$ and $f(\theta)$ outside the integral sign. These quantities vary only slightly over the area concerned.

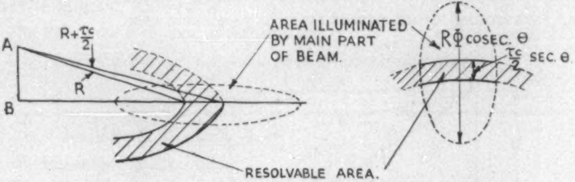

A
B
$R+\frac{\tau c}{2}$
R
AREA ILLUMINATED BY MAIN PART OF BEAM.
RΦ COSEC. θ
$\frac{\tau c}{2}$ SEC. θ
RESOLVABLE AREA.

Figure 2. Resolvable area well within main area of illumination.

The integral in (7) may be evaluated from the polar diagram of the radar, but often a sufficiently good approximation is

$$\int F^4(\theta,\phi)\,d\phi = \Phi, \qquad \ldots\ldots(8)$$

$$P_r = \frac{PGA}{16\pi^2 R^3}\tau\mathbf{c}\tan\theta f(\theta)\,.\,\Phi. \qquad \ldots\ldots(9)$$

This result is valid only as long as $\frac{\tau\mathbf{c}}{2}\sec\theta \ll R\Phi\operatorname{cosec}\theta$, which is so for sufficiently small angles of depression.

§ 2.2.

The mathematical condition for this state is

$$R\Phi\operatorname{cosec}\theta \ll \frac{\tau\mathbf{c}}{2}\sec\theta.$$

We express ds as a function of both $d\theta$ and $d\phi$:

$$ds = R^2\,d\theta\,d\phi\operatorname{cosec}\theta. \qquad \ldots\ldots(10)$$

$$\therefore\quad dP_r = \frac{PGA}{8\pi^2 R^2} f(\theta) F^4(\theta,\phi)\,d\theta\,d\phi. \qquad \ldots\ldots(11)$$

We again assume that $F^4(\theta, \phi)$ is a function of θ and ϕ which varies much more rapidly than R and $f(\theta)$.

$$\therefore \quad P_r = \frac{PGA}{8\pi^2 R^2} f(\theta) \iint F^4(\theta, \phi) d\theta\, d\phi. \qquad \ldots\ldots(12)$$

This double integral is evaluated from a knowledge of the polar diagram, and is approximately constant as long as

$$R\Phi \operatorname{cosec} \theta \ll \frac{\tau c}{2} \sec \theta ,$$

i.e. for sufficiently large angles of depression.

Thus it can be seen that for the echo power received from the surface there are two formulae, each applicable over a different region. If θ_0 is the value of θ for which $R\Phi \operatorname{cosec} \theta = \frac{\tau c}{2} \sec \theta$, then

when $\theta \ll \theta_0$, equation (7) is valid;

when $\theta \gg \theta_0$, equation (12) is valid.

In both cases $f(\theta)$ has appeared as a multiplying factor, and from the experimental results this important function may be easily calculated.

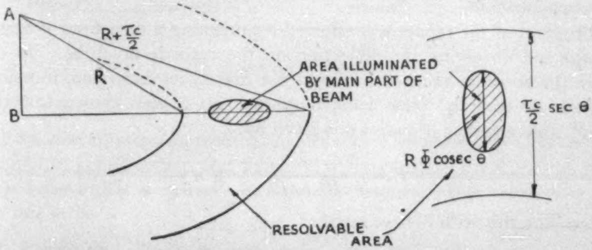


Figure 3. Main area of illumination well within resolvable area.

In the theoretical approach to the problem, the object was to determine mathematically the coefficient $f(\theta)$ in terms of θ, λ, and roughness of the sea. The starting point of the analysis was the assumption that a sea wave is sinusoidal in section. The model of the sea was two-dimensional, the waves being cylindrical.

The return from a single wave was first calculated, and then this echo averaged over all waves. The statistical distribution for the amplitudes of the waves was taken to be

$$p(y) = \frac{y}{A^2} \exp\left(\frac{-y^2}{A^2}\right) \qquad y>0, \qquad \ldots\ldots(13)$$

where A is the mean amplitude.

Similarly the distribution assumed for the wave-lengths of the sea waves was

$$q(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\Delta)^2}{2\sigma^2}\right\}, \qquad \ldots\ldots(14)$$

where Δ is the mean wave-length and σ is the standard deviation.

On the basis of these assumptions it has been shown * that $f(\theta)$ should be independent of θ for angles greater than some critical angle, and should behave as θ^n for very small angles. We shall see later that these conclusions are not altogether borne out by the results.

We should not expect the mathematical treatment to give completely correct results, since besides the assumptions already mentioned, no account has been taken of "white horses", which undeniably destroy the sinusoidal shape of the sea waves.

§3. EXPERIMENTAL DETAILS

The results were obtained from three separate sets of experiments which will be dealt with in order of importance.

§3.1. Using four radar equipments, a systematic investigation into sea clutter was conducted by Telecommunications Research Establishment at Great Orme's Head, Llandudno, 500 ft. above sea-level, with a wide angle of view straight out to sea. Two of the radars, K3 and K4, operated on a wave-length of 1·25 cm., one, X3, on a wave-length of 3 cm., and one, S1, on a wave-length of 10 cm. X3 had a conical beam of width 4°·5, a peak power output of 30 kw., and a pulse length of 0·5 μsec. K4 had a beam of width 0°·58 in azimuth and 15° in elevation, a peak power of 20 kw., and could send out pulses of either 0·5 μsec. or 0·2 μsec. duration.

Calibration of the radars was effected by injecting a measured voltage at the I.F. stage and observing the deflection on the cathode-ray tube. In order to measure the absolute value of $f(\theta)$, and not merely its variations, measurements were made on the echo from a metal sphere. It is well known that the echo power P_s from a sphere of radius a is given by

$$P_s = \frac{PG}{4\pi R^2}\,\frac{\pi a^2 A}{4\pi R^2} \qquad \text{if } a \gg \lambda. \qquad \ldots\ldots(16)$$

If we compare this with (9), we see that

$$\frac{P_r}{P_s} = \frac{R\tau\mathbf{c}\tan\theta\,.\,f(\theta)\,.\,\Phi}{\pi a^2}$$

$$= \frac{h\tau\mathbf{c}\Phi f(\theta)}{\pi a^2}. \qquad \ldots\ldots(17)$$

Thus $f(\theta)$ may be calculated without the measurement of the rather awkward quantities P, G, A, etc.

The procedure adopted for measuring the sea echo at range R was to tilt the beam until maximum signal was obtained at range R on the tube. In this case the range of angles investigated was $\frac{1}{2}° < \theta < 5°$, a region well within the condition $\theta < \theta_0$. Formulae (7) and (9) are applicable to all these results.

§3.2. The second set of experiments was conducted in an aircraft and the range of angles covered was $15° < \theta < 90°$. Only an X Band Set was used, with a beam width of 6°, a pulse length of 1 μsec. and a peak power output of 20 kw. Unfortunately no measurements on a sphere were made, and so only variations

---

* Unpublished paper by G. G. Macfarlane.

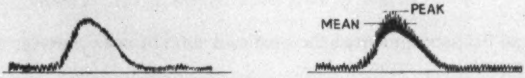
in $f(\theta)$ could be observed. However, some idea of the absolute value of $f(\theta)$ can be obtained by linking up the results of this set with those for X Band at Llandudno.

§ 3.3. A number of observations on S Band have been made by the Army Operational Research Group at Ashford Flank. The range covered is small, being $1^\circ < \theta < 5^\circ$, when equation (7) is valid. From these measurements $f(\theta)$ has been calculated.

§ 4. THE RESULTS

§ 4.1. *The shape of the echo*

If the resolvable area is very large, as is the case for a high-flying aircraft with a radar beam width of say 6° and a pulse length of 1 μsec., then the echo appearing on the cathode-ray tube looks like figure 4 (*a*). If the resolvable area is rather smaller, then the appearance of the echo is as shown in figure 4 (*b*). The latter case was observed at Llandudno. Two levels were discernible which we shall call the peak and mean sea levels, the mean sea level being about 6 db. down on the peak sea. If the resolvable area is made very small, then ultimately individual waves will be resolved and the echo will consist of a large number of separate " blips ".



(*a*) (*b*)

Figure 4. The appearance of the echo on "A" scope.

The appearance of the echo

We have concentrated on reading the peak level when two levels have existed, and although it is not certain that this level corresponds to the P_r of § 2, there is reason to believe that it behaves in a similar manner. Our calculation of $f(\theta)$ refers to this level.

§4.2. *Results on* $\lambda = 1{\cdot}25$ *cm.*

Over the limited range for which results are available, i.e. $\frac{1}{2}^\circ < \theta < 5^\circ$, the results have shown that the echo power P_r varies inversely as the fourth power of the range. Figure 5 is an example, typical of many such observations.

Since $h = R \operatorname{cosec} \theta \approx R \cot \theta$,

$$P_r \propto \frac{1}{R^3} \tan \theta,$$

from which it is clear on referring to equation (9) that $f(\theta)$ is independent of θ in the interval $\frac{1}{2}^\circ < \theta < 5^\circ$.

Although $f(\theta)$ is independent of θ over the interval $\frac{1}{2}^\circ < \theta < 5^\circ$, it is not independent of the state of the sea. As a criterion of the degree of roughness of the sea, it was thought best to make observations on the wave-height. This was done from a trawler and a submarine operating in the vicinity. The values of $f(\theta)$ for different seas are given in figure 6, from which it can be seen that $f(\theta)$ does not increase perceptibly after a wave-height of 3 feet. Even though no detailed readings are available for wave-heights of the order of 6 to 12 inches, it is known that the echo decreases very rapidly in this region. The maximum value of $10 \log f(\theta)$ observed on $\lambda = 1\cdot25$ cm. was -12.

On the equipment K4 it was possible to verify the reasoning which leads to the echo power being proportional to the pulse length. Decreasing τ from 0·5 μsec. to 0·2 μsec. produced the predicted drop in echo power to within the limits of experimental error.

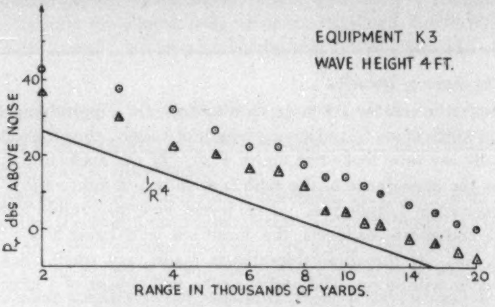


Figure 5. The dependence of echo power on range for $\lambda = 1\cdot25$ cm.

§ 4.3. *Results on wave-length of* 3 *cm.*

As indicated previously, there are two sources of information on $\lambda = 3$ cm. The range $\frac{1}{2}^\circ < \theta < 5^\circ$ is covered by the Llandudno experiments, with which we shall deal first.

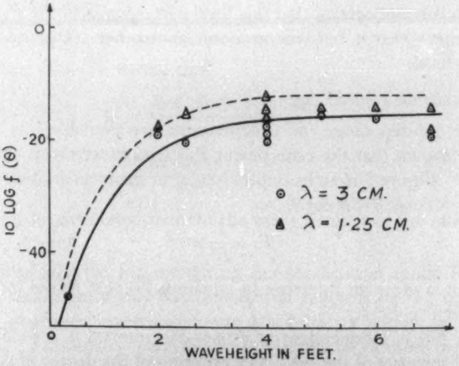


Figure 6. The dependence of $f(\theta)$ on the state of the sea.

The results are very similar to those on 1·25 cm. wave-length. Over the interval $\frac{1}{2}^\circ < \theta < 5^\circ$ the echo was found to follow the $1/R^4$ law, implying that $f(\theta)$ is independent of θ. A very large number of results confirms this; a typical example is shown in figure 7.

For the dependence of $f(\theta)$ on the state of the sea the reader should consult figure 6, where $10 \log f(\theta)$ over the interval $\frac{1}{2}° < \theta < 5°$ is plotted against wave-height. Three feet appears to be the height above which $f(\theta)$ does not vary with the height of the waves. The maximum observed value of $10 \log f(\theta)$ was $-15{\cdot}5$.

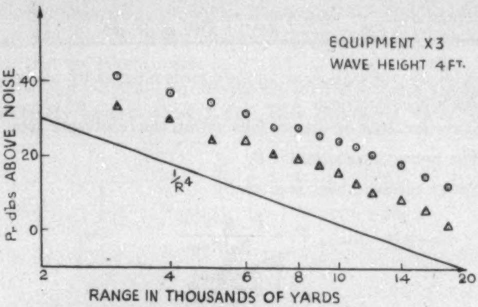


Figure 7. The dependence of echo power on range for $\lambda = 3$ cm.

From the second source of information on 3 cm. wave-length, the aircraft results, we have been able to calculate $f(\theta)$ over the whole range $15° < \theta < 90°$. The variation with θ for one particular state of the sea is shown in figure 8.

The constancy of $f(\theta)$ found at Llandudno appears to extend up to 40°. As before, the level of AB, the flat portion of the curve, depends on the roughness of the sea, but no wave-height measurements are available and so there is no check on the data in figure 6. The aircraft results indicate that, although the level AB depends on the state of the sea, $10 \log f(90°)$ is independent of this variable and the value of $10 \log f(90°)$ is $+5$. This means that the echo from the sea at normal incidence is 5 db. greater than the echo obtained from a perfectly rough surface in the same direction.

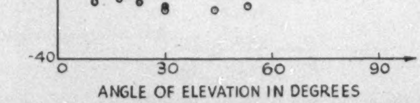


Figure 8. The dependence of $f(\theta)$ on θ, the angle of elevation ($\lambda = 3$ cm.).

Let us compare the echo received back from a perfectly smooth conducting surface with the echo obtained from the sea surface, both being viewed at normal incidence on 3 cm.

For the sea surface,

$$P_r = \frac{PGA}{8\pi^2 R^2} f(\theta) \iint F^4(\theta, \phi)\, d\theta \,.\, d\phi,$$

where $10 \log f(\theta) = 5$; or we may avoid the double integral by using the following reasoning, which is less accurate.

All the power incident on the sea falls within the resolvable area.

∴ The power intercepted is P.

∴ Power received back is given by

$$P_r = \frac{P \,.\, f(\theta) G \lambda^2}{2\pi R^2 4\pi}$$

$$= \frac{PG\lambda^2 f(\theta)}{8\pi^2 R^2}.$$

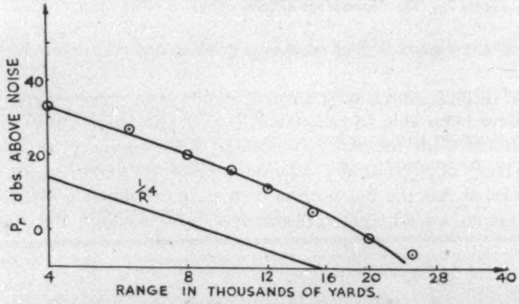


Figure 9. The dependence of echo power on range for $\lambda = 10$ cm.

The power P_r' received from a perfectly smooth surface is the power received by the radar if it were situated at the image point

$$P_r' = \frac{PGA}{4\pi(2R)^2} = \frac{PG^2\lambda^2}{64\pi^2 R^2}.$$

$$\therefore \quad \frac{P_r}{R_r'} = \frac{8f(\theta)}{G} \approx \frac{25}{G}.$$

Therefore, if the aerial gain is greater than 25, more energy is received back from a perfectly smooth surface than from the sea.

§ 4.4. *Results on wave-length of* 10 *cm.*

The relatively few results on 10 cm. wave-length indicate that the echo power varies inversely as the fourth power of the range for angles of depression from 5° down to 1°. Thereafter the decrease is more rapid. This is clearly shown in figure 9.

Thus, for the interval $1^\circ < \theta < 5^\circ$, $f(\theta)$ is independent of θ. Its dependence upon the state of the sea in this region of angles is shown in figure 10. Unfortunately, measurements of wave-height were not made, and it is doubtful whether the use of wind-speed as a criterion of the degree of roughness is justifiable. Some attempt has been made to replace wind-speed by wave-height. The high degree of scatter in the results is undoubtedly due to this bad criterion.

It appears that on 10 cm. wave-length $f(\theta)$ behaves somewhat the same as did $f(\theta)$ on 3 cm. and 1·25 cm., but the maximum value is only reached at a wave-height considerably greater than 3 ft. This maximum value of $10 \log f(\theta)$ is -21.

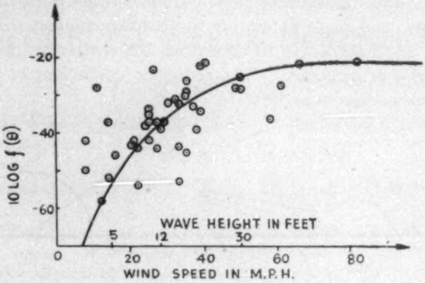


Figure 10. The dependence of $f(\theta)$ on the wind velocity ($\lambda = 10$ cm.).

§ 4.5. *The dependence of $f(\theta)$ on λ*

We have seen that over a considerable range of angles $f(\theta)$ is independent of θ, but dependent on the state of the sea. Its value increases as the sea gets rougher, tending to a maximum value which depends on the wave-length of the radar. This dependence is clear from the accompanying table.

Wave-length	Max. value of $10 \log f(\theta)$
1·25 cm.	-12
3 cm.	$-15{\cdot}5$
10 cm.	-21

This leads immediately to the result

$$f(\theta)_{\max} = \frac{\lambda_0}{\lambda},$$

where $\lambda_0 = 1/12$ cm. and λ is expressed in cm.

It might be thought that $f(\theta)$ would tend to the same value on all wave-lengths as the wave-heights increase, but this is obviously not the case, anyway not for small angles. This fact suggests that the greatest echo is obtained from ripples and not from the large waves, and that although the wave-height increases to very large values, the quantity and size of ripple do not increase after a certain wave-height has been achieved.

§ 5. CONCLUSIONS

Although the results so far obtained are not inconsiderable, there is much room for expansion in the subject. Only on 3 cm. wave-length has the whole range of angles of depression been investigated. The same information on K and S bands would not be without advantage. Then the scattering coefficient $f(\theta, \lambda$, state of sea) would be known for all θ, for $\lambda = 1$ to 10 cm., and for all seas from calm to very rough. When this coefficient is known, the echo obtained from the sea with any radar equipment may be calculated.

Another possible extension is the examination of the shape of the scattering polar diagram. Up till now, only the echo in the direction of incidence, i.e. the radar echo, has been investigated. With a receiver completely detached from the transmitter, it would be possible to investigate the energy scattered in other directions, and so to determine the scattering polar diagram. It would be interesting and informative to determine the scattering polar diagram for a number of angles of incidence. The reflection coefficient of the surface might be measured at the same time.

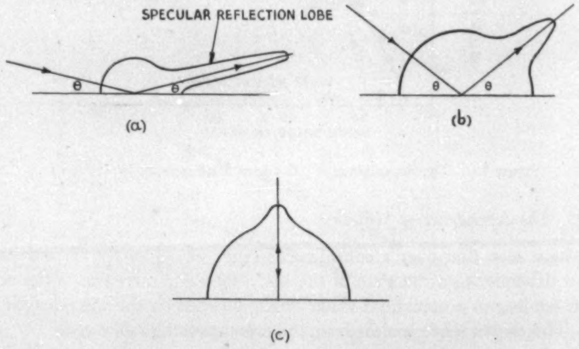


Figure 11. Hypothetical scattering polar diagram.

It is anticipated that the section of the scattering polar diagram in the plane of incidence will be similar to those shown in figure 11. For small angles of incidence the reflection lobe will be very large compared with the scattered energy. As the angle increases, the reflection lobe moves as shown and also becomes smaller. Thus more energy is scattered than in figure 11 (*a*). Finally, at normal incidence, the lobe of specular reflection is barely significant, the return being only 5 db. above that from a " perfectly rough " surface. However, the diagrams given are purely hypothetical, and experiments will have to be made before any reliance at all can be placed upon them.

§ 6. ACKNOWLEDGMENTS

The work described in this paper was carried out at the Telecommunications Research Establishment, Malvern, as part of the war-time programme of research of the Ministry of Aircraft Production, and is now published by permission of the Ministry of Supply (Air). All the diagrams are Crown Copyright.

Thanks are due to J. R. Atkinson, A. L. Cockcroft, E. R. Wiblin, and G. J. R. MacLusky, who were responsible for the experimental work; to R. A. Finlayson, who took part in the earlier analysis of the results; to H. G. Hooker for helpful criticism throughout; to the Superintendent, Army Operational Research Group, for certain results at $\lambda = 10$ cm.; and finally, to the Royal Navy for its helpful co-operation during the experiments.

EXTENSION OF GRIFFITH'S THEORY OF RUPTURE TO THREE DIMENSIONS

By R. A. SACK,

H. H. Wills Physical Laboratory, University of Bristol

(Now at Department of Applied Mathematics, University of Liverpool)

Communicated by N. F. Mott, F.R.S.; MS. received 29 *May* 1946

ABSTRACT. Griffith's theory of rupture of brittle materials is extended to materials containing circular cracks. It is found that (*a*) the tensile strength of brittle material in one direction is not affected by stresses at right angles to it; (*b*) the result differs from Griffith's by a factor depending on Poisson's ratio of the material and lying between 1·57 and 1·81.

§ 1. INTRODUCTION

A THEORY of rupture of brittle solids has been developed by Griffith (1921). This theory assumes the existence of a large number of small cracks on the surface or in the interior of the solid, rupture being primarily conditioned by the extension of a crack already existing, and not by the formation of a new one. The spreading of a crack is accompanied by an increase in energy proportional to the increase of the surface; on the other hand, the internal surface increases the elastic deformability of the material, thus leading to a lower free energy when the material is subject to given external stresses. It depends on the magnitudes and directions of crack and stresses whether the total contribution is positive or negative; if an increase of the size of a crack leads to a diminution of the free energy, the system will become unstable, the crack will spread and the material be fractured.

The quantitative development of Griffith's theory has to take account of the size and shape of the crack and of its orientation relative to the applied stress. For mathematical reasons, Griffith (1921, 1924) has confined the exact theoretical treatment to the problem in two dimensions, i.e. that of a plane plate with an elliptic cavity going right through it. By considering the crack as an ellipse of vanishing minor axis and assuming that Hooke's law holds good up to the corners of the crack, he finds that rupture will occur when the stress P normal to a crack reaches the critical value

$$P_c = \sqrt{\left(\frac{2ET}{\pi c}\right)} \quad \text{for plane stress,} \qquad \ldots\ldots(1)$$

$$P_c = \sqrt{\left(\frac{2ET}{\pi c(1-\sigma^2)}\right)} \quad \text{for plane strain,*} \qquad \ldots\ldots(2)$$

where E denotes Young's modulus of the material, σ Poisson's ratio, T the surface tension, and c the width of the crack if situated on the surface of the plate or half this width if situated in the interior.

The origin of the Griffith cracks has not been properly explained, but for some crystals their existence has been demonstrated experimentally (Andrade and Martindale, 1936). If they are situated on the surface, their length is very likely to be much greater than their width c, so that the two-dimensional model is a good approximation. Glass and many crystals are materials of this type; their tensile strength can be varied considerably by suitable treatment of their surface (Joffe and Levitzky, 1925; Reinkober, 1937). If, however, the cracks are situated in the interior of the material, there is no reason to expect that their length will greatly exceed their width, and a three-dimensional model has to be used.

It is the purpose of this paper to provide the required three-dimensional theory. The conditions of rupture are calculated for a brittle amorphous solid containing a number of plane circular cracks when one of the principal stresses acts normally to the plane of one of these cracks. This normal principal stress must be tensile, for if it is compressive the faces of the crack will be pressed together and can exert tractions on each other, whereas the calculations are based on the assumption that the internal surface is free from tractions. The influence of other cracks can be neglected as the additional stresses and strains set up by them fall off rapidly with distance.

In § 2 the strain correction is calculated, which is set up by the crack if the material is subject to a homogeneous stress †; in § 3 the diminution of the free energy of the system due to the crack is determined, and the tensile strength is calculated in § 4. The case of oblique fracture connected with compressive external stresses will not be considered in this paper.

Since the present calculations were first carried out, I. N. Sneddon (1945) has calculated independently the stresses set up by a circular crack subject to an internal pressure. Instead of the curvilinear coordinates employed in the present paper, Sneddon uses cylindrical polars; he arrives at the displacements and stresses by the method of Hankel transforms developed by Harding and himself (1945). The advantage of Sneddon's method is that the stresses are obtained in well-determined directions; against this must be set the drawback that the evaluation of the stresses involves the use of rather lengthy integrals. The final result for the tensile strength is the same for both methods.

§ 2. CALCULATION OF STRAIN CORRECTION

To calculate the strain correction set up by the crack a method due to Neuber (1934) will be used. If $x_i (i=1, 2, 3)$ are the Cartesian coordinates of a point within the elastic solid and q_i the displacements from the unstrained state in the direction of the three axes, Hooke's law can be expressed by the equation

---

* The formula $\sqrt{[2ET(1-\sigma^2)/\pi c]}$ for plane strain as stated by Griffith (1924) is obviously erroneous.

† This calculation forms a special case of a problem treated by Neuber in his dissertation (Munich, 1932); this, however, has not been available to the writer.

$$S_{ik}=G\left[\frac{\partial q_i}{\partial x_k}+\frac{\partial q_k}{\partial x_i}+\delta_{ik}\frac{2\sigma}{1-2\sigma}\operatorname{div}\boldsymbol{q}\right],\quad i,k=1,2,3,\qquad \ldots\ldots(3)$$

where the $\boldsymbol{S}_{ik}$ denote the stress components in rectangular coordinates, G is the modulus of rigidity, and

$$\delta_{ik}=\begin{cases}1 \\ 0\end{cases}\text{if}\quad \begin{matrix} i=k \\ i\neq k\end{matrix},$$

$$\operatorname{div}\boldsymbol{q}=\frac{\partial q_1}{\partial x_1}+\frac{\partial q_2}{\partial x_2}+\frac{\partial q_3}{\partial x_3}.\qquad \ldots\ldots(4)$$

If no external volume forces are acting, the condition for equilibrium is

$$\sum_k\frac{\partial S_{ik}}{\partial x_k}=0,\qquad i=1,2,3,\qquad \ldots\ldots(5)$$

which with (3) leads to

$$\Delta q_i+\frac{1}{1-2\sigma}\frac{\partial}{\partial x_i}\operatorname{div}\boldsymbol{q}=0,\qquad i=1,2,3.\qquad \ldots\ldots(6)$$

According to Neuber the general solution of (6) is given by

$$q_i=-\frac{\partial F}{\partial x_i}+2\alpha V_i,\qquad i=1,2,3,\qquad \ldots\ldots(7)$$

where

$$F=V_0+x_1V_1+x_2V_2+x_3V_3,\qquad \ldots\ldots(8)$$

$$\alpha=2(1-\sigma)\qquad \ldots\ldots(9)$$

and the V_k are harmonic functions satisfying

$$\Delta V_k=0,\qquad k=0,1,2,3.\qquad \ldots\ldots(10)$$

In the expressions (7) and (8) one of the functions $V_k(k\neq 0)$ can be put equal to zero without loss of generality.

If, instead of Cartesians, curvilinear orthogonal coordinates u_ν are used, such that an element of length ds at any point is given by

$$ds^2=\sum_\nu h_\nu^2 du_\nu^2\qquad \ldots\ldots(11)$$

(Greek suffixes λ, μ, ν refer to curvilinear coordinates), where the $\boldsymbol{h}_\nu$ are functions of the $\boldsymbol{u}_\nu$, then the displacements q_ν have to be taken in the direction of increasing u_ν and the stresses $\boldsymbol{S}_{\mu\nu}$ in the corresponding planes. The relation (3) has now to be replaced by

$$\frac{S_{\mu\nu}}{G}=\frac{h_\mu}{h}\frac{\partial}{\partial u_\nu}\left(\frac{q_\mu}{h_\mu}\right)+\frac{h_\nu}{h_\mu}\frac{\partial}{\partial u_\mu}\left(\frac{q_\nu}{h_\nu}\right)+\delta_{\mu\nu}\left[\frac{2-\alpha}{\alpha-1}\operatorname{div}\boldsymbol{q}+\frac{2}{h_\mu}(\boldsymbol{q}\cdot\operatorname{grad}h_\nu)\right],\qquad \ldots\ldots(12)$$

where

$$\operatorname{div}\boldsymbol{q}=\frac{1}{h_\lambda h^\mu h}\left[\frac{\partial}{\partial u_\lambda}(h_\mu h_\nu q_\lambda)+\ldots\right],\qquad \ldots\ldots(13)$$

whereas the Laplacian Δ is expressed by

$$\Delta=\frac{1}{h_\lambda h_\mu h_\nu}\left[\frac{\partial}{\partial u}\left(\frac{h_\mu h_\nu}{h_\lambda}\frac{\partial}{\partial u_\lambda}\right)+\ldots\right].\qquad \ldots\ldots(14)$$

(The dots denote cyclic permutation of the suffixes λ, μ, ν.)

Though the functions V_k can be expressed in terms of the new coordinates u_ν, it is not as a rule possible to take them along directions that vary from point to point; they still have to be taken along the Cartesian axes. As

$$\cos(x_k, u_\nu) = \frac{1}{h_\nu}\frac{\partial x_k}{\partial u_\nu}, \qquad \ldots\ldots(15)$$

the solution (7) becomes

$$q_\nu = \frac{1}{h_\nu}\left(-\frac{\partial F}{\partial u_\nu} + 2\alpha \sum_k \frac{\partial x_k}{\partial u_\nu} V_k\right), \qquad \ldots\ldots(16)$$

where F is still defined by (8).

For problems relating to a plane circular disc or crack of infinitely small thickness, the appropriate coordinates are oblate spheroidal coordinates ψ, ϑ, ϕ, where the surfaces

$\psi = \text{const.}$ represent oblate spheroids,
$\vartheta = \text{const.}$,, hyperboloids of revolution of one sheet,
$\phi = \text{const.}$,, planes through the x_1-axis.

The Cartesians are expressed in the new coordinates as

$$\left.\begin{aligned} x_1 &= c \sinh\psi \cos\vartheta, \\ x_2 &= c \cosh\psi \sin\vartheta \cos\phi, \\ x_3 &= c \cosh\psi \sin\vartheta \sin\phi. \end{aligned}\right\} \qquad \ldots\ldots(17)$$

The surface $\psi = 0$ corresponds to the circular disc $x_1 = 0$, $x_2^2 + x_3^2 \leqslant c^2$; each point of the circle is represented twice, corresponding to the two sides of the disc. The functions h_ν are derived from (17),

$$\left.\begin{aligned} h_\psi^2 = h_\vartheta^2 &= c^2(\sinh^2\psi + \cos^2\vartheta) = c^2h^2, \\ h_\phi &= c \cosh\psi \sin\vartheta. \end{aligned}\right\} \qquad \ldots\ldots(18)$$

Thus the Laplacian becomes

$$\Delta = \frac{1}{c^2h^2 \cosh\psi \sin\vartheta}\left[\frac{\partial}{\partial\psi} \cosh\psi \sin\vartheta \frac{\partial}{\partial\psi} + \frac{\partial}{\partial\vartheta} \cosh\psi \sin\vartheta \frac{\partial}{\partial\vartheta}\right] + \frac{1}{c^2 \cosh^2\psi \sin^2\vartheta}\frac{\partial^2}{\partial\phi^2}. \qquad \ldots\ldots(19)$$

The normal solutions of the equation $\Delta V = 0$ are

$$\left.\begin{aligned} &\Pi_n{}^m(\sinh\psi) P_n{}^m(\cos\vartheta) e^{im\phi} \\ \text{and} \quad &T_n{}^m(\sinh\psi) P_n{}^m(\cos\vartheta) e^{im\phi}, \end{aligned}\right\} \qquad \ldots\ldots(20)$$

where the functions $P_n{}^m$ are the usual Legendre polynomials, whereas $\Pi_n{}^m(x)$ and $T_n{}^m(x)$ represent the Legendre functions of the first and second kind respectively of the argument ix (cf. e.g. Hobson, 1931), or explicitly

$$\left.\begin{aligned} P_n{}^m(x) &= \text{const.}\,.\,(1 - x^2)^{m/2} \frac{d^{n+m}}{dx^{n+m}}(1 - x^2)^n, \\ \Pi_n{}^m(x) &= \text{const.}\,.\,(1 + x^2)^{m/2} \frac{d^{n+m}}{dx^{n+m}}(1 + x^2)^n, \\ T_n{}^m(x) &= \text{const.}\,.\,\Pi_n{}^m(x) \int_x^\infty \frac{dx}{[\Pi_n{}^m(x)]^2(1 + x^2)}. \end{aligned}\right\} \qquad \ldots\ldots(21)$$

If one of the principal stresses P of the external system is normal to the crack, there will be no shear stresses $S_{\psi\vartheta}$ and $S_{\psi\phi}$ acting across its faces, neither will the other two principal stresses be affected by the presence of the crack, but the normal component $S_{11}=S_{\psi\psi}=P$ must be compensated for as there should be no tractions exerted on the inner surface. This disturbance will exhibit axial symmetry and fall off with increasing distance. Thus the only terms to be considered are those containing $T_n{}^0$; terms with $m\neq 0$ or $\Pi_n{}^0$ need not be considered in the calculations.

The first three functions T are

$$\left.\begin{aligned} &T_0(\sinh\psi)=T_0{}^0(\sinh\psi)=\cot^{-1}(\sinh\psi),\\ &T_1(\sinh\psi)=T_1{}^0(\sinh\psi)=T_0(\sinh\psi)\sinh\psi-1,\\ &T_2(\sinh\psi)=T_2{}^0(\sinh\psi)=T_0(\sinh\psi)(3\sinh^2\psi+1)-3\sinh\psi. \end{aligned}\right\}\quad\ldots\ldots(22)$$

If in (7), (8) and (16) we put $V_2=V_3=0$ the displacements q_ψ and q_ϑ are

$$\left.\begin{aligned} q_\varphi&=\frac{1}{h}\left[-\frac{1}{c}\frac{\partial V_0}{\partial\psi}+(2\alpha-1)\cosh\psi\cos\vartheta V_1-\sinh\psi\cos\vartheta\frac{\partial V_1}{\partial\psi}\right],\\ q_\vartheta&=\frac{1}{h}\left[-\frac{1}{c}\frac{\partial V_0}{\partial\vartheta}-(2\alpha-1)\sinh\psi\sin\vartheta V_1-\sinh\psi\cos\vartheta\frac{\partial V_1}{\partial\vartheta}\right]. \end{aligned}\right\}\quad\ldots\ldots(23)$$

According to (12) the stresses acting across the faces $\psi=0$ are given by

$$\left.\begin{aligned} \frac{S_{\psi\psi}}{2G}\bigg|_{\psi=0}&=\frac{1}{c^2\cos^2\vartheta}\left[-\frac{\partial^2V_0}{\partial\psi^2}+\tan\vartheta\frac{\partial V_0}{\partial\vartheta}+\alpha c\cos\vartheta\frac{\partial V_1}{\partial\psi}\right]_{\psi=0},\\ \frac{S_{\psi\vartheta}}{2G}\bigg|_{\psi=0}&=\frac{1}{c^2\cos^2\vartheta}\left[-\frac{\partial^2V_0}{\partial\psi\partial\vartheta}-\tan\vartheta\frac{\partial V_0}{\partial\psi}+(\alpha-1)c\cos\vartheta\frac{\partial V_1}{\partial\vartheta}\right]_{\psi=0}. \end{aligned}\right\}$$

$$\ldots\ldots(24)$$

The required conditions

$$S_{\psi\vartheta}=0,\quad S_{\psi\psi}=\text{const.}\quad\text{for }\psi=0\qquad\ldots\ldots(25)$$

are satisfied by a disturbance specified by the harmonic functions V_0 and V_1:

$$\left.\begin{aligned} &V_0=\tfrac{1}{3}(\alpha-1)c^2A[T_0(\sinh\psi)+\tfrac{1}{2}T_2(\sinh\psi)(\tfrac{3}{2}\cos^2\vartheta-\tfrac{1}{2})],\\ &V_1=cAT_1(\sinh\psi)\cos\vartheta. \end{aligned}\right\}\quad\ldots\ldots(26)$$

Introduction of these functions into (24) yields

$$S_{\psi\vartheta}=0,\quad S_{\psi\psi}=\pi GA,\quad\text{for }\psi=0.\qquad\ldots\ldots(27)$$

At large distances the spheroidal coordinates approximate spherical polars with

$$c\sinh\psi=r\qquad\ldots\ldots(28)$$

and ϑ and ϕ in their usual meanings. The functions T become

$$T_0(\sinh\psi)=\frac{c}{r}-\frac{c^3}{3r^3}+\ldots,\quad T_1(\sinh\psi)=-\frac{c^2}{3r^2}+\ldots,\qquad\ldots\ldots(29)$$

whereas T_2 can be neglected as it falls off in a higher power. The functions h_ν are now

$$h_r=1,\quad h_\vartheta=r,\quad h_\phi=r\sin\vartheta,\qquad\ldots\ldots(30)$$

so that according to (12) the stresses are now expressed as

$$S_{rr}=G\left[2\frac{\partial q_r}{\partial r}+\frac{2-\alpha}{\alpha-1}\operatorname{div}q\right],\quad S_{r\vartheta}=G\left[\frac{\partial q_\vartheta}{\partial r}-\frac{q_\vartheta}{r}+\frac{1}{r}\frac{\partial q_r}{\partial\vartheta}\right].\qquad\ldots\ldots(31)$$

The deformations at large distances will be, according to (16), (26), (29) and (30),

$$\left.\begin{aligned} q_r &= \frac{c^3A}{9r^2}[-(4-\alpha)-(4\alpha+2)(\tfrac{3}{2}\cos^2\vartheta-\tfrac{1}{2})]+\dots, \\ q &= \frac{2c^3A}{3r^2}(\alpha-1)\cos\vartheta\sin\vartheta+\dots, \end{aligned}\right\} \qquad \dots\dots(32)$$

from which the stresses are derived by means of (31):

$$\left.\begin{aligned} \frac{S_{rr}}{2G} &= \frac{2Ac^3}{9r^3}[4-\alpha+(8+\alpha)(\tfrac{3}{2}\cos^2\vartheta-\tfrac{1}{2})]+\dots, \\ \frac{S_{r\vartheta}}{2G} &= \frac{Ac^3}{3r^3}(4-\alpha)\sin\vartheta\cos\vartheta+\dots \end{aligned}\right\} \qquad \dots\dots(33)$$

The dots in (29), (32) and (33) denote terms to a higher power in c/r, which can be neglected.

§ 3. ENERGY CORRECTION

After determining the additional stresses and deformations set up by the crack, we have to consider how the elastic energy of the stressed material is affected by such a disturbance. If the material has the shape of a large sphere of radius R_0 containing a circular crack, it would not be correct simply to take the solution for the uncracked sphere and to superpose on this the strains (23), (26), (32). The condition of equilibrium demands that the tractions across the external surface given by (12) should agree accurately with the imposed stresses. This condition can be satisfied by putting

$$\left.\begin{aligned} q_r &= \frac{c^3A}{9r^2}[-(4-\alpha)-(4\alpha+2)P_2]+\frac{2(\alpha-1)}{4-\alpha}Br+CrP_2-(2-\alpha)Dr^3P_2, \\ q_\vartheta &= \left[\frac{2c^3A}{3r^2}(\alpha-1)-\tfrac{3}{2}Cr+(\alpha+\tfrac{3}{2})Dr^3\right]\sin\vartheta\cos\vartheta, \end{aligned}\right\} \qquad \dots\dots(34)$$

where

$$P_2=P_2{}^0(\cos\vartheta)=\tfrac{3}{2}\cos^2\vartheta-\tfrac{1}{2}, \qquad \dots\dots(35)$$

and by virtue of (31)

$$\left.\begin{aligned} \frac{S_{rr}}{2G} &= \frac{2Ac^3}{9r^3}[(4-\alpha)+(8+\alpha)P_2]+B+CP_2+\left(1-\frac{\alpha}{2}\right)Dr^2P_2, \\ \frac{S_{r\vartheta}}{2G} &= \left[\frac{4-\alpha}{3}\frac{c^3A}{r^3}-\tfrac{3}{2}C+\frac{2-\alpha}{2}Dr^2\right]\sin\vartheta\cos\vartheta. \end{aligned}\right\} \qquad \dots\dots(36)$$

The B term corresponds to a pure hydrostatic pressure, the term with C to a combination of pure shearing stresses; the D term is a correction with the property

$$\operatorname{div} \boldsymbol{q}=7(\alpha-1)Dr^2P_2. \qquad \dots\dots(37)$$

At the crack only terms to zeroth power in c/R will have to be considered, the D term will be negligible and the vanishing of the tractions across the faces is expressed as

$$S_{11}=S_{\psi\psi}=2G(\tfrac{1}{2}\pi A+B+C)=0. \qquad \dots\dots(38)$$

The free energy of an elastic solid under the influence of stresses is equal to half the integral of the product of the surface tractions into the displacements; this integral has to be taken over the whole surface and to be given a negative sign (cf. Griffith, 1921).

In the present case, in view of the axial symmetry, this integral will be

$$-2W=2\pi\int q_r S_{rr} R_0^2 \sin\vartheta\, d\vartheta+2\pi\int q_\vartheta S_{r\vartheta} R_0^2 \sin\vartheta\, d\vartheta. \qquad \ldots\ldots(39)$$

As R_0 tends towards infinity, terms of W depending on a negative power of R_0 will vanish. This is the reason why in (32) and (34) higher powers in $1/R_0$ than the second have been neglected, and in (33) and (36) higher powers than the third.

If a pure hydrostatic pressure N is exerted, (36) and (38) lead to

$$\left.\begin{aligned} \frac{N}{2G}&=B+\tfrac{2}{9}(4-\alpha)\frac{c^3}{R_0^3}A,\\ 0&=\tfrac{1}{2}\pi A+B. \end{aligned}\right\} \qquad \ldots\ldots(40)$$

Hence

$$\left.\begin{aligned} B&=\frac{N}{2G}\left[1+\frac{4}{9\pi}(4-\alpha)\frac{c^3}{R_0^3}\right],\\ A&=-\frac{N}{\pi G}. \end{aligned}\right\} \qquad \ldots\ldots(41)$$

In view of the orthogonality of P_2 and $P_0(\cos\vartheta)=1$, the values of C and D do not enter into the result. The equation (39) becomes, with the use of (34), (36) and (41),

$$-2W=2\pi R_0^2\frac{N^2}{2G}\int_0^\pi\left[\frac{2(\alpha-1)}{4-\alpha}\left(1+\frac{4c^3}{9\pi}\,\frac{4-\alpha}{R_0^3}\right)R_0+\frac{2c^3}{9\pi}\,\frac{4-\alpha}{R_0^2}\right]\sin\vartheta\, d\vartheta, \qquad \ldots\ldots(42)$$

$$W=-\frac{\pi N^2}{G}\left[2\frac{(\alpha-1)}{4-\alpha}R_0^3+\frac{2\alpha}{3\pi}c^3\right]. \qquad \ldots\ldots(43)$$

In the absence of the crack the energy would have been

$$W_0=-\frac{2\pi}{G}\,\frac{\alpha-1}{4-\alpha}N^2R_0^3. \qquad \ldots\ldots(44)$$

The introduction of the crack therefore increases the free energy of the system by an amount

$$W_1=W-W_0=-\frac{2\alpha}{3G}N^2c^3. \qquad \ldots\ldots(45)$$

Similarly, if an external stress $(M,\ -\tfrac{1}{2}M,\ -\tfrac{1}{2}M)$ is applied, the tractions S_{rr} and $S_{r\vartheta}$ for $r=R_0$ will be

$$S_{rr}=MP_2(\cos\vartheta), \quad S_{r\vartheta}=-\tfrac{3}{2}M\sin\vartheta\cos\vartheta. \qquad \ldots\ldots(46)$$

The equality of the stresses (36) and (46), in conjunction with (38), leads to

$$W_1=-\frac{2\alpha}{3G}M^2c^3. \qquad \ldots\ldots(47)$$

A combination of these two types of stresses yields by the same method

$$W_1=-\frac{2\alpha}{3G}(N+M)^2c^3. \qquad \ldots\ldots(48)$$

Finally an additional stress of the type $(0,\ L,\ -L)$ will not affect the value of W_1.

§ 4. RESULT AND DISCUSSION

If a brittle body is acted upon by a uniform stress $(P,\ Q,\ R)$, (48) shows that the presence of a circular crack of radius c in a plane normal to P alters the free energy of the solid by an amount

$$W_1 = -\frac{2\alpha}{3G}P^2c^3 = -\frac{8(1-\sigma^2)}{3E}P^2c^3, \qquad \ldots\ldots(49)$$

irrespective of Q and R, provided P is tensile and c is appreciably smaller than the dimensions of the solid. On the other hand, the crack has a surface energy amounting to

$$W_2 = 2\pi c^2 T, \qquad \ldots\ldots(50)$$

where T represents the surface tension of the material. Thus the total free energy contribution due to the crack is

$$\Delta W = W_1 + W_2 = 2\pi c^2 T - \frac{8(1-\sigma^2)}{3E}P^2c^3. \qquad \ldots\ldots(51)$$

For varying c this expression has a maximum when

$$c = \frac{\pi ET}{2P^2(1-\sigma^2)}. \qquad \ldots\ldots(52)$$

If c is greater than this value, or alternatively if

$$P > P_c = \sqrt{\left(\frac{\pi ET}{2c(1-\sigma^2)}\right)}, \qquad \ldots\ldots(53)$$

the crack will become unstable and spread. If plane circular cracks of radius c are distributed at random in a brittle solid, rupture will be determined solely by the maximum tensile stress; the tensile strength in one direction will not be affected by the presence of smaller tensile (and, within limits, compressive) stresses at right angles to it.

Comparison of the result (53) with the plane case (1) and (2) shows that the tensile strengths differ by a factor ranging between $\frac{1}{2}\pi = 1{\cdot}57$ and $\pi/\sqrt{3} = 1{\cdot}81$. If the cracks are supposed to be plane ellipses, then for a given minor axis c the variation of the major axis from infinity to c affects the tensile strength of the material by that factor only. The tensile strength is thus primarily determined by the length of the minor axes of the cracks and must lie between the limits given by Griffith (1921 and 1924) and in equation (53).

ACKNOWLEDGMENTS

I have to thank Professor N. F. Mott, F.R.S., for suggesting this problem, and for valuable discussions in the course of the calculations. I have also to thank the Chief Scientific Officer, Ministry of Supply, for permission to publish this paper.

REFERENCES

Andrade and Martindale, 1936. *Phil. Trans.*, **235**, 69.
Griffith, 1921. *Phil. Trans.*, **221**, 180.
Griffith, 1924. *II Int. Congress Appl. Maths., Delft*, p 55.
Harding and Sneddon, 1945. *Proc. Camb. Phil. Soc.*, **41**, 16.
Hobson, 1931. *Theory of Spherical and Ellipsoidal Harmonics* (Cambridge : The University Press).
Joffe and Levitzky, 1925. *Z. Phys.*, **31**, 567.
Neuber, 1934. *Z. Ang. Math. Mech.*, **14**, 203.
Reinkober, 1937. *Phys. Z.*, **37**, 112.
Sneddon, 1946. *Proc. Roy. Soc.*, A (in the press).

THE THERMODYNAMIC PROPERTIES OF THE PRODUCTS OF HIGH-PRESSURE COMBUSTION

By J. CORNER,
Armament Research Department, Ministry of Supply

MS. received 27 *March* 1946

ABSTRACT. The internal energy, heat content, specific heats and equilibrium constants of the normal products of combustion are expanded as power series in the density. At pressures up to 40 tons/sq. in., corresponding to densities of about 0·35 gm./c.c., only the first three terms need be kept. Thus the pressure-dependence of the thermodynamic properties can be evaluated from a knowledge of the second and third virial coefficients of the various product gases. Tables are presented based on the most recent intermolecular forces, and covering the range from 1600° K. to 4000° K. These tables have found considerable application to internal ballistics, and give "covolumes" of propellants with a systematic error of less than 5%.

---

§1. INTRODUCTION

From the equation of state of a gas we can derive the contributions of gas-imperfection to thermodynamic properties such as its internal energy. In the application to the products of high-pressure combustion we desire to find the energy and heat content of the products of explosion, at temperatures from 1500 to 4000° K. and at densities up to 0·35 gm./c.c. Under these conditions it is sufficient to retain the second and third virial coefficients. Theoretical expressions for these are known in the case of the pure gases, and the arbitrary parameters can be chosen to fit the observed second virial coefficient at low temperatures. Hence we can find the equations of state of the pure gases at high temperatures, and so, by using thermodynamic formulae, we can find their internal energies and heat contents. Relating the virial coefficients of the mixture to those of the constituents, by means of a simple combination rule, we derive, finally, the properties of the product mixture and also the pressure-correction to the equilibrium constants which determine the composition.

The details of this programme are given in the present paper. A similar aim, with largely the same methods, has been pursued by Hirschfelder and his co-workers, in work which was given a restricted circulation while the present work was being carried out. The points of difference have been considered sufficiently important to warrant the completion of the present work. These differences lie chiefly in the treatment of carbon dioxide, and the extent to which one may regard the second virial coefficient as independent of temperature. There are also minor differences in the equations and parameters used.

The basis of the present work is the expression of the second virial coefficient of a gas as a simple function of the parameters of the intermolecular field. There is a general expression for the virial coefficient in the form of an integral over the intermolecular field at all distances; the difficulty has always been in the discovery of an intermolecular potential of plausible form, with an integrable second virial coefficient. For molecules with spherical symmetry, we use a special case of the well-known Lennard-Jones potential (Lennard-Jones, 1924). This has been used in the discussion of nitrogen, carbon monoxide and hydrogen (§§ 4 and 5 of this paper). Steam is covered by Stockmayer's recent development (Stockmayer, 1941) of a potential which represents the field of a polar molecule with a short-range repulsion of spherical symmetry. We have also taken into account the work of Hirschfelder, McClure and Weeks (1942) and Margenau and Myers (1944) on steam. Carbon dioxide is an example of a cigar-shaped, non-polar molecule, whose virial coefficient has been studied by Corner (1946).

The most obvious application of our tables is to problems of internal ballistics, particularly to the testing of propellants in closed vessels. The standard thermochemical tables used in closed-vessel work in Britain are those of Pike (unpublished). The specific heats and equilibrium constants tabulated there refer to zero pressure, and the explosion temperatures, pressures and product compositions calculated from these tables are therefore incorrect. Of these quantities only the pressure can be measured experimentally; the theoretical pressure is adjusted by the introduction of a "covolume", chosen to give the right pressure at normal loading densities. The theoretical temperatures and compositions are in error by unknown amounts.

The tables in the present paper are intended to be used in conjunction with tables such as those of Pike. They purport to give real pressures of explosion, and this is confirmed by the comparison with experimental results in § 12. To put it another way, we may say that these tables enable covolumes to be predicted. The tables also give true temperatures and compositions, both of which are difficult to test by experiment. At the same time this fact makes theoretical results all the more important in physical and chemical problems of internal ballistics, such as questions of erosion by heat transfer and chemical action.

§ 2. BASIS OF THE TABLES

We wish to find the internal energy and heat content of the products of combustion at pressures up to 40 tons/sq. in. and at temperatures of order 1500–4000° K.* The values at zero density can be deduced from spectroscopic data. The pressure-corrections can be derived from the equations of state of the products, by thermodynamic formulae. This use of the equation of state to obtain theoretical specific heats is now common in chemical engineering, but we have been able to find only one published application at the temperatures encountered in internal ballistics. Burlot (1932) has calculated the specific heats of CO_2 and N_2 from an empirical equation of state, claimed to hold up to 100° C. and 300 atmospheres. This is far too slender a basis for even qualitative results under ballistic conditions. The only other early work which bears on this subject is that of Crow and Grimshaw (1931), who attempted to derive molecular radii from dielectric constants. When suitably corrected the results fitted the observed covolumes of non-flashless propellants. This has no more than an empirical significance, for in fact the "covolume" may arise not only from the volume occupied by the molecules but also from the resulting change in the specific heats; the latter cannot be calculated from the rough equation of state used by Crow and Grimshaw. It is true that this change of the explosion temperature has been proved, from the present tables, to be negligible, but this is a result which could not have been deduced from an approximate equation of state.

---

* The upper limit is settled by the explosion temperature of the hottest Service propellants, about 3700° K. ; temperatures below 1500° K. cause difficulties with terms neglected in our formulae, and, in any case, the assumption of chemical equilibrium, necessary to the use of thermo-chemical tables, is violated in many processes of combustion ending at lower temperatures.

The calculation of heat and energy contents requires an accurate equation of state. Experimental results do not go beyond 700° K., so there is a wide range to be covered by extrapolation before we reach the region 1500–4000° K., of interest in combustion. The extrapolation is, in fact, sufficiently wide to be unbridgeable in an empirical way. Fortunately there is now a considerable body of theory to help us, and it plays an essential part in our work. It is perhaps as well to emphasize that this theory has been verified and tested on the equations of state of gases at normal temperatures, so that, in effect, the theory is being used as an extrapolation formula with the backing of statistical mechanics. The work is straightforward, and the only doubtful point is whether our equations of state and specific heats are, in fact, sufficiently accurate. We shall describe various tests which have been applied. Meanwhile we may say that the basic data (equations of state of the pure gases up to 600–700° K.) and the theory are well established, and future modifications are likely to be small; also there seems to be little hope of working back from experiments at ballistic temperatures. Accordingly our results are not likely to be much altered by new experimental data at low temperatures and not likely to be superseded by a more direct method.

In a purely empirical extrapolation to high temperatures the only reason for confidence in the results would be the accuracy of fit to experiments at normal temperatures and possibly the successful prediction of covolumes. The essential basis of our theoretical extrapolation is the reasonable nature of the intermolecular forces assumed. For example, we treat N_2, CO and H_2 as having fields of spherical symmetry, for this assumption is not only sufficient to account for the low-temperature equation of state but is also reasonably consistent with what we know of their molecular structure. The same applies to H_2O, except that here we have used a field which is effectively that of a dipole with a short-range repulsion of spherical symmetry; a theory of such intermolecular fields was produced by Stockmayer in 1941. The other main product of propellant explosions is CO_2, which presented the only real difficulty in our work. This molecule has a cigar shape, and it is not plausible to use a field of spherical symmetry even though this gives a good fit to experiments at normal temperatures. We have therefore devised a molecular field which is, roughly speaking, of cigar or cylindrical shape, and we have found the equation of state of a gas of such molecules (Corner, 1946). This equation we have fitted to observations on CO_2 up to 300° C., to find the parameters of the potential for this molecule. From this point the calculation of heat and energy contents at high temperatures has followed the same course as for the simpler molecules.

It would be possible to apply the same methods to N_2, CO and H_2; the gain in accuracy is probably too small to justify at present the time that would be required.

The fact that the products of explosion are not perfect gases has an effect on the equilibrium constants which determine the composition of the products. This effect, too, has been calculated from the equation of state by familiar thermodynamic formulae.

§ 3. EQUATION OF STATE OF A PURE GAS AND ITS THERMODYNAMIC PROPERTIES

Let P atmospheres be the pressure, V c.c. the volume, and T° K. the temperature of a mole of pure gas. Up to the state of a saturated vapour the pressure can be expressed by the series

$$\frac{PV}{RT}=1+\frac{B}{V}+\frac{C}{V^2}+\frac{D}{V^3}+\dots\dots,$$

where the coefficients B, C, D, etc. are functions only of T and the nature of the gas. R is 82·06 (Birge, 1941). B and C are the second and third virial coefficients respectively. We shall use

$$\frac{PV}{RT}=1+\frac{B}{V}+\frac{C}{V^2}, \qquad \dots\dots(1)$$

which is sufficiently accurate for our purposes. For example, at these temperatures B for nitrogen is of the order 30 c.c./mole, C of order 150 (c.c./mole)^2, so that at densities around 0·35 gm./c.c. the successive terms on the right of (1) are 1, 0·4 and 0·02. It is obvious that even at this density, which lies at the upper end of the range occurring in combustion problems, the neglect of higher terms in the expansion leads to errors of only 0·1% in the pressure.

It may be worth pointing out that the three-term power-series in (1) is considerably more accurate * than the series

$$\frac{PV}{RT}=1+R(T)P+S(T)P^2,$$

which is often used in chemical engineering. The latter equation has the form $V=f(P, T)$, which is a virtue in that field. In ballistics the most convenient form is $P=f(V, T)$, which is fortunately a property of the more accurate equation (1).

Let $E(T, V)$ and $H(T, V)$ be the internal energy and the heat content of the mole of gas at temperature T° K. and volume V c.c. In general,

$$\left(\frac{\partial E}{\partial V}\right)_T=T\left(\frac{\partial P}{\partial T}\right)_V-P,$$

and using (1) we find

$$E(T, V)-E(T,\infty)=-RT^2\left(\frac{B'}{V}+\frac{C'}{2V^2}\right), \qquad \dots\dots(2)$$

where dashes denote derivatives with respect to T. We write this as

$$E(T, V)=E_0(T)+\frac{E_1(T)}{V}+\frac{E_2(T)}{V^2}, \qquad \dots\dots(3)$$

where

$$\left.\begin{aligned} E_1&=-RT^2B' \\ \text{and}\quad E_2&=-\tfrac{1}{2}RT^2C'. \end{aligned}\right\} \qquad \dots\dots(4)$$

---

* Cf. p. 203 of Beattie and Stockmayer (1940).

From $H=E+PV$ we have

$$H(T, V)=H_0(T)+\frac{H_1}{V}+\frac{H_2}{V^2}, \qquad \ldots\ldots(5)$$

where

$$\left.\begin{aligned} H_1&=\mathbf{R}T(B-TB') \\ \text{and} \qquad H_2&=\mathbf{R}T(C-\tfrac{1}{2}TC'). \end{aligned}\right\} \qquad \ldots\ldots(6)$$

This can be checked by the relation

$$\left(\frac{\partial H}{\partial P}\right)_T=V-T\left(\frac{\partial V}{\partial T}\right)_P.$$

Let A be the Helmholtz free energy of the system. Since $\left(\frac{\partial A}{\partial V}\right)_T=-P$ we find from (1),

$$A(T, V)=A(T)-\mathbf{R}T\ln V+\frac{\mathbf{R}TB}{V}+\frac{\mathbf{R}TC}{2V^2}. \qquad \ldots\ldots(7)$$

Let C_V and C_P be the specific heats at constant volume and pressure respectively. From $\left(\frac{\partial C_V}{\partial V}\right)_T=T\left(\frac{\partial^2 P}{\partial T^2}\right)_V$ it follows that

$$C_V(T, V)=C_{V,0}(T)+\frac{C_{V,1}}{V}+\frac{C_{V,2}}{V^2}, \qquad \ldots\ldots(8)$$

where

$$C_{V,1}=-\mathbf{R}T(2B'+TB''), \qquad \ldots\ldots(9)$$

$$C_{V,2}=-\tfrac{1}{2}\mathbf{R}T(2C'+TC''). \qquad \ldots\ldots(10)$$

Also

$$\begin{aligned} C_p-C_V&=T\left(\frac{\partial P}{\partial T}\right)_V\left(\frac{\partial V}{\partial T}\right)_P \\ &=\mathbf{R}\left\{1+\frac{\alpha}{V}+\frac{\beta}{V^2}\right\}^2\left\{1+\frac{2B}{V}+\frac{3C}{V^2}\right\}^{-1}, \qquad \ldots\ldots(11) \end{aligned}$$

where

$$\alpha=B+TB', \qquad \ldots\ldots(12)$$

$$\beta=C+TC'. \qquad \ldots\ldots(13)$$

We shall use these results when we have obtained theoretical formulae for B and C. To this we now proceed, dealing in turn with each of the major products of combustion.

§4. HYDROGEN

Consider a pair of similar molecules, whose intermolecular potential energy W depends only on r, the distance between their centres. Let the zero of W be taken at infinite separation of the molecules. It can be shown that in classical statistical mechanics

$$B=2\pi\mathbf{N}\int_0^\infty\left[1-\exp\left\{-\frac{W}{\mathbf{k}T}\right\}\right]r^2dr, \qquad \ldots\ldots(14)$$

where $\mathbf{N}$ is Avogadro's number and $\mathbf{k}$ is Boltzmann's constant.

Lennard-Jones (1924) has integrated this as a convergent series when

$$W(r)=\phi\left(\frac{mp}{m-p}\right)\left[\frac{1}{m}\left(\frac{r_0}{r}\right)^m-\frac{1}{p}\left(\frac{r_0}{r}\right)^p\right] \qquad ......(15)$$

and m and p are constants satisfying $m>p>3$; r_0 and ϕ are parameters characteristic of the molecule, being respectively the position and depth of the minimum of the potential. The behaviour of many simple molecules is well represented by the special case

$$W=\phi\left[\left(\frac{r_0}{r}\right)^{12}-2\left(\frac{r_0}{r}\right)^6\right] \qquad ......(16)$$

which we shall use for H_2, N_2 and CO. This potential is often referred to as the Lennard-Jones (12, 6) potential.

In the literature there are three sets of (ϕ, r_0) for hydrogen. The earliest is that of Lennard-Jones (1931), deduced from observed second virial coefficients; Hirschfelder and Roseveare (1939) have derived ϕ and r_0 from a study of the Beattie-Bridgeman equation. Both these analyses neglected quantum corrections, which are appreciable for hydrogen right up to room temperature. Allowance for these effects has been made by de Boer and Michels (1938). Table 1 compares the second virial coefficient at high temperatures from these three sets of parameters. The column marked " Keyes " was computed from a semi-empirical equation which was proposed by Keyes (1941) as a summary of the experimental data. For extrapolation to these high temperatures it is less reliable than the results of the full theoretical analysis, to which the semi-empirical equation is an approximation. All four sets of B show the same trend with temperature and good agreement in magnitude. We have used the most recent theoretical values, those of de Boer and Michels, in the preparation of our tables. These parameters are

$$\phi/\boldsymbol{k}=37^\circ\cdot02;\ r_0=3\cdot28\ \text{A}. \qquad ......(17)$$

Table 1. B (c.c./mole) of H_2 at high temperatures

Temp. (°K.)	L–J	H, R	B, M	Keyes
1500	16·2	13·5	16·4	15·7
2000	15·7	13·1	16·0	15·1
2500	15·2	12·7	15·6	14·7
3000	14·8	12·4	15·2	14·2
3500	14·4	12·1	14·8	13·8
4000	14·1	11·9	14·5	13·5

We have omitted all quantum corrections to the second virial coefficient. These corrections are largest for light molecules at low temperatures, and in the worst case within our range (hydrogen at 1500° K.) they rise to 1%. For the other products the corrections are very small; for example, they amount to only 1 part in 2000 for nitrogen at 1500° K.

We come now to the third virial coefficient, C. Recently $C(T)$ has been integrated for the intermolecular potential (16) by de Boer and Michels (1939) and Montroll and Mayer (1941). We define a new dimensionless variable by

$$\tau = \boldsymbol{k}T/\phi. \qquad \ldots\ldots(18)$$

It can be shown that

$$C = \left\{\frac{\sqrt{2}\pi N r_0}{3}\right\}^2 C^*(\tau) \qquad \ldots\ldots(19)$$

merely by dimensional analysis. Here C^* is a function of τ only. de Boer and Michels computed C^* for τ between 4 and 1·5 by numerical integration of a function which itself involved a numerical double integral; Montroll and Mayer started with the same formula for C^*, but transformed it into numerical integration of a function which has been computed by a single integration. We have plotted the two sets of results in figure 1. There is some disagreement between the two sets, especially at small values of $1/\tau$, which is just where we intend to use C^*. However, the Montroll-Mayer curve is the smoother, and a comparison with the rather scattered experimental values of C^* is more favourable to Montroll and Mayer's curve. We shall assume their curve, and shall use it to extrapolate C^* towards $1/\tau = 0$, where $C^* = 0$. Various series of integral and fractional powers of $1/\tau$ have been used to represent the behaviour of C^*, but none of these differs significantly from

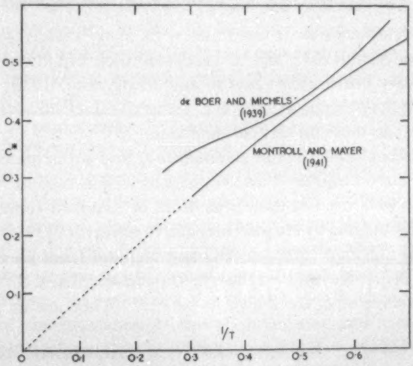


Figure 1. Alternative theoretical results for the coefficient C^*.

$$C^* = \frac{0{\cdot}9315}{\tau} - \frac{0{\cdot}0738}{\tau^2} + \frac{0{\cdot}212}{\tau^3} \qquad \ldots\ldots(20)$$

in the region of τ up to 40. This simple formula fits the curve of Montroll and Mayer to within the errors of reading from the latter, for τ greater than 1·5 (cf. figure 1). We have used (17), (18), (19) and (20) to get C and dC/dT for hydrogen at temperatures from 1500 to 4000° K. This corresponds to the range of τ from 40 to 100, where C is extremely small, of order 15 (c.c./mole)^2; B is of order 15 c.c./mole.

Table 2 gives some idea of the size of the corrections caused by gas imperfections in hydrogen under typical ballistic conditions. These results were calculated from the tables given in § 11 of the present paper, combined with those compiled by Dr. H. H. M. Pike for the ideal gas.

Table 2. Gas imperfection in hydrogen at 3000° K. and 100 c.c./mole

	Ideal gas	From our tables	Percent difference
Pressure (atmos.)	2460	2839	15·4
$H-H$ (300° K., zero density), cal./mole	23500	24540	4·4
$E-E$ (300° K., zero density), cal./mole	17540	17670	0·8
C_V, cal./deg. mole	6·808	6·866	0·9
C_P, cal./deg. mole	8·795	8·813	0·2

§ 5. NITROGEN AND CARBON MONOXIDE

These two substances have molecules with many closely similar properties (cf. Grimm and Wolf, 1933), and in particular their equations of state for the gas phase are very much alike (Beattie and Bridgeman, 1928; Newitt, 1940). We shall treat both substances together, using the data on nitrogen, which are more extensive than those on carbon monoxide.

The most recent values of the parameters r_0 and ϕ for nitrogen are those of Corner (1940), based on Joule-Thomson data as well as second virial coefficients. The latter alone were used in the earlier work of Lennard-Jones (1931). Other values have been obtained by Hirschfelder and Roseveare (1939) from the constants of the Beattie-Bridgeman equation (CO and N_2) and from an analysis of Joule-Thomson data (N_2). Keyes (1941) has given a semi-empirical correlative equation. The resulting values of B are shown in table 3. They are in good agreement, and the most recent theoretical values lie near the mean of the others. We have, therefore, used the Corner parameters

Table 3. B (c.c./mole) of N_2 at high temperatures

Temp. (°K.)	Keyes	L–J	JC	H, R
1500	31·1	33·4	31·9	32·3
2000	31·5	34·3	32·6	33·1
2500	31·4	34·4	32·7	33·3
3000	31·0	34·3	32·6	33·2
3500	30·6	34·0	32·3	32·9
4000	30·2	33·7	32·0	32·6

$$\phi/\boldsymbol{k}=95^\circ{\cdot}3_3;\ r_0=4{\cdot}10\ \text{A}. \qquad \ldots\ldots(21)$$

The computation of B, C, E and H followed the same course as for hydrogen. Table 4 shows the magnitude of the effects for nitrogen.

Table 4. Gas imperfection in nitrogen at 3000° K. and 100 c.c./mole

	Ideal gas	From our tables	Percent difference
Pressure (atmos.)	2460	3290	33·7
$H-H$ (300° K., zero density), cal./mole	22170	24290	9·6
$E-E$ (300° K., zero density), cal./mole	16800	16920	0·7
C_V, cal./deg. mole	6·876	7·018	2·1
C_P, cal./deg. mole	8·863	9·048	2·1

§ 6. STEAM

This cannot be treated by the field of spherical symmetry assumed for nitrogen and hydrogen. Stockmayer (1941) has found the second virial coefficient for a potential which represents a dipole interaction superposed on the usual London attraction and short-range repulsion; both the latter terms were taken to be spherically symmetrical. He has applied the theory to the experimental data for steam, assuming the repulsive potential to have the form Ar^{-24} This is a much higher index than the values, around 12, found for simple non-polar molecules. Hirschfelder, McClure and Weeks (1942) have repeated the comparison between experiment and the theory with the more plausible index of 12. We have compared the virial coefficients at 3200° K. from their parameters and those of Stockmayer; B is respectively 7·75 and 12·3 c.c./mole, TdB/dT is 9·1 and 11·6 c.c./mole, and T^2d^2B/dT^2 is $-24·6$ and $-31·1$ c.c./mole. In computing our tables we have used the parameters found by Hirschfelder, McClure and Weeks.

Margenau and Myers (1944) have produced an intermolecular field for steam which they claim is more plausible than that of Stockmayer. We have therefore repeated all our calculations on steam, using the Margenau-Myers field, and have computed covolumes of propellants (cf. § 12) from both sets of tables. The Margenau-Myers field increases the error in the covolume from 4 to 6%. We have followed the indication of this empirical test, and in the tables of § 11 the values for steam are based on the work of Hirschfelder, McClure and Weeks.

There has been no theoretical investigation of the third virial coefficient for polar molecules. To estimate C we proceeded thus: the *theoretical* second virial coefficient from 1400 to 4000° K. was analysed in the usual way with the Lennard-Jones (12, 6) potential (16), and it was found that the virial coefficient could best be represented by the virial from (16) if the constants in the latter were given the values

$$\phi/\boldsymbol{k}=560°; \quad \frac{\sqrt{2}\pi \boldsymbol{N}r_0^3}{3}=26·1 \text{ c.c./mole.} \qquad \ldots\ldots(22)$$

These are plausible values.* We have assumed that from 1400 to 4000° K. the third virial coefficient also is approximately the same as that from the potential (16) with the constants (22).

---

* For non-polar molecules the critical point is given by

$$V_c \simeq 1·48\left(\frac{\sqrt{2}\pi Nr_0^3}{3}\right) \text{ and } T_c \simeq 1·3\phi/k,$$

and in conjunction with (22) they would give $T_c = 750°$ K. and $V_c = 39$ c.c. mole. Observed values for water are 647° K. and 45 c.c./mole. This shows that the parameters (22) are of the right order of magnitude.

Table 5 shows the size of the corrections to thermodynamic data for steam.

Table 5. Gas imperfection in steam at 3000° K. and 100 c.c./mole

	Ideal gas	From our tables	Percent difference
Pressure (atmos.)	2460	2664	8·3
$H-H$ (300° K., zero density), cal./mole	29930	29860	−0·2
$E-E$ (300° K., zero density), cal./mole	24570	24000	−2·3
C_V (cal./deg. mole)	11·107	11·246	1·3
C_P (cal./deg. mole)	13·094	13·566	3·6

§ 7. CARBON DIOXIDE

This is another molecule whose intermolecular field cannot be represented by terms with spherical symmetry. Both long and short-range forces depend on the mutual orientation of the pair of molecules considered. There does not appear to be any theoretical discussion of such a case in the literature. The requirement is the second virial coefficient for molecules which are non-polar, roughly cylindrical, and stiff.

It is obvious that molecules which resemble cylinders form a very wide class, for not only the nuclear framework but also the nature of the nuclei in different parts of the molecule vary from one case to another. The intermolecular field, therefore, has to satisfy conflicting requirements of generality and accuracy of representation, as well as the essential point that it must give an integrable expression for the second virial coefficient. The latter requirement is not important in the case of spherical symmetry, since the integral has in that case only one variable of integration; numerical integration is always possible, though it may be laborious. When the molecules have lower symmetry the second virial coefficient is given by an integral with at least four variables of integration, while the specification of the field itself requires at least one more parameter than in the case of spherical symmetry. Numerical integration is therefore impracticable.

We shall explain in outline how we have tried to find an intermolecular potential which satisfies the three conditions of accuracy, generality and integrability. Details are being published elsewhere (Corner, 1946). As a first step it was assumed that at any fixed relative orientation the intermolecular potential depended on the distance between the centres of the molecules in a way which has been found to give a good representation of the behaviour of molecules with spherical symmetry. The second virial coefficient was thus reduced to a triple integral. This potential has two parameters, R_0 and ϕ_0, which are respectively the position and depth of the minimum potential. For cylindrical molecules ϕ_0 and R_0 are functions of the relative orientation. In the next step it was necessary to find the dependence of $\phi_0^{\frac{1}{2}}$ and R_0^3 on the three angles which specify the orientation and on the length/breadth ratio of the molecule. The first model investigated was a pair of centres of force separated by a certain distance; the potential assumed for the interaction of such centres of force was the usual potential for saturated spherical groups. For various orientations of two such molecules the mutual potential was evaluated numerically and the values of ϕ_0 and R_0 found by interpolation. This process was repeated with molecules represented by three and four centres of force distributed along the central axis. The three-centre model gave results differing considerably from the two-centre type, because the latter has a " waist " which cannot be expected as a feature of a real molecule. The four-centre type showed little difference in behaviour from the three-centre type. It was concluded that the four-centre model gave an adequate representation of how the normal intermolecular potential is altered by a change from spherical to cylindrical form. This four-centre model was then investigated for various " molecular lengths ". Formulae were found which fitted the variation of $\phi_0^{\frac{1}{2}}$ and R_0^3 with the three orientation variables and the length/diameter ratio of the model. Finally these formulae were substituted in the integral for the second virial coefficient and this was integrated by expansion in series.

It is thought that the potential energy so derived is likely to be a good approximation for most molecules of cylindrical type, while, of course, it may be considerably in error for special cases. The molecule is assumed to be of roughly the same nature at both ends, and there are certainly many molecules which do not satisfy this condition. It is, however, true for carbon dioxide, and the other conditions of a fairly stiff nuclear framework and zero resultant dipole moment are also satisfied in this case.

In this way we have computed tables which give the second virial coefficient in terms of three variables: r_0 and ϕ are related to the depth and position of the minimum potential at a standard orientation; the third variable is effectively the length of the nuclear framework of the molecule. As this length tends to zero, r_0 and ϕ become the parameters already defined for spherically symmetrical systems in § 4. We have also computed tables for the Joule-Thomson coefficient at low pressure, μ_0; the connection between this and the second virial coefficient was pointed out by Whitelaw (1934), and has been applied to the determination of intermolecular forces by Hirschfelder, Ewell and Roebuck (1938).

In applying these tables to carbon dioxide we took the length of the nuclear framework to be 2·32 A. (Adel and Denison, 1935). The most recent virial data are those of Michels and Michels (1935), from 0 to 150° C. They agree to within 4 c.c./mole with the virial coefficients deduced from Amagat's data (1893) by Bridgeman (1927). The Joule-Thomson coefficient has been found by Roebuck, Murrell and Miller (1942) over a wide range of pressure and from -50 to 300° C. We have extrapolated to zero pressure the values at temperatures over 0° C.; below this temperature the extrapolation is rather uncertain.

The specific heat at zero density, $C_P{}^\circ$, which was needed in the analysis, was based on Kassel's work (Kassel, 1934). The best values of ϕ and r_0 were found to be

$$\phi/\boldsymbol{k} = 198^\circ; \quad \frac{\sqrt{2}\pi N r_0^3}{3} = 110 \text{ c.c.} \qquad \ldots\ldots(23)$$

This potential gives second virial coefficients with errors of order 2 c.c./mole, i.e. about 2% at room temperatures; the error in $\mu_0 C_P{}^\circ$ rises to about 10 c.c./mole at low temperatures, around 0° C., but is of order 5 c.c./mole (2%) over most of the temperature-range. Quantum corrections were included in this comparison.

At high temperatures the second virial coefficient becomes the same as that of the spherically symmetrical potential (16) with the same values of ϕ and r_0; this was arranged by the definitions of these parameters. For carbon dioxide the difference between the two virial coefficients is less than 1% at 1400° K. In preparing our tables we have therefore neglected the correction to the virial coefficient of the spherical potential. Hence we may compare the set (23) with others which have been derived from fields with spherical symmetry. Hirschfelder and Roseveare (1939) found (175°, 131 c.c.) from the Beattie-Bridgeman constants for carbon dioxide. Hirschfelder, McClure, Curtiss and Osborne (unpublished) have derived (185°, 120 c.c.) from isotherms of Michels, Blaisse and Michels (1937) (at medium rather than low pressures) and have checked these values against the Joule-Thomson data of Roebuck, Murrell and Miller (1942).

This spherically-symmetrical potential gives as good a fit to the low-temperature gas data as our potential with cylindrical symmetry, but we think the latter gives a more reliable extrapolation to high temperatures. Table 6 shows some high-temperature virial coefficients for these potentials.

Table 6. Second virial coefficients of CO_2 at 3000° K.

Potential	B (c.c./mole)
(23)	56·2
Hirschfelder and Roseveare	60·3
(185°, 120 c.c.)	62·1

Since, at high temperatures, the theoretical second virial coefficient approximates to that of the Lennard-Jones (12, 6) potential, it is reasonable to assume that the third virial coefficient also can be represented by the results of Montroll and Mayer, calculated from a (12, 6) potential. We have therefore used (18), (19) and (20), with the constants (23), in finding the third virial.

Table 7. Gas imperfection for CO_2 at 3000° K. and 100 c.c./mole

	Ideal gas	From our tables	Percent difference
Pressure (atmos.)	2460	4025	63·6
$H-H$ (300° K., zero density), cal./mole	36710	40240	9·6
$E-E$ (300° K., zero density), cal./mole	31350	31080	−0·9
C_V (cal./deg. mole)	13·020	13·277	2·0
C_P (cal./deg. mole)	15·007	15·567	3·7

§8. COMBINATION RULES

The discussion of §3 referred to a pure gas. For a mixture at low densities it is equally permissible to write

$$\frac{PV}{RT}=1+\frac{B_1}{V}+\frac{C_1}{V^2}, \qquad \ldots\ldots(24)$$

where V is the volume occupied by the gas of $\boldsymbol{N}$ molecules, $\boldsymbol{N}$ being Avogadro's number; B_1 and C_1 are functions of temperature and composition. The internal energy and other thermodynamic data are still given by equations similar to those in § 3, with the differentiations carried out at constant composition. The zero of internal energy, for example, refers to *the same* gases at zero density and 300° K.; C_P is the specific heat at constant pressure, *keeping the composition fixed.*

Coming now to the mixture of product gases, we consider the products of 1 gm., namely $[CO_2]$ moles of CO_2, [CO] of CO, etc. Let

$$[CO_2]+[CO]+\ldots\ldots=n, \qquad \ldots\ldots(25)$$

and let V c.c. be the volume occupied by these gases. Then the molar volume of the mixture is (V/n) c.c., which is the volume to be used in equation (24). Hence

$$\frac{PV}{n\boldsymbol{R}T}=1+\frac{nB_1}{V}+\left(\frac{n}{V}\right)^2 C_1. \qquad \ldots\ldots(26)$$

We must now obtain B_1 from the second virial coefficients of the pure gases. Several such combination rules are in use (cf. a review by Beattie and Stockmayer, 1940), but for our purposes only the simplest can be considered. This is the linear sum rule:

$$nB_1=[CO_2]B(CO_2)+[CO]B(CO)+\ldots. \qquad \ldots\ldots(27)$$

We use also the analogous formula:

$$nC_1=[CO_2]C(CO_2)+[CO]C(CO)+\ldots. \qquad \ldots\ldots(28)$$

From these and the formulae of § 3, we arrive at

$$E(T,V)=\Sigma[CO]\left\{E_{CO,0}(T)+\frac{n}{V}E_{CO,1}(T)+\left(\frac{n}{V}\right)^2 E_{CO,2}(T)\right\} \qquad \ldots\ldots(29)$$

for the internal energy relative to the same gases at 300° K. and zero density. $E_{CO,0}(T)$ is the internal energy of a mole of CO at zero density and temperature T; $E_{CO,1}$ and $E_{CO,2}$ are the corrections from the second and third virial coefficients. The summation is over all constituents of the mixture. The explicit formulae for $E_{CO,1}$ and $E_{CO,2}$, given in equation (4), are

$$E_{CO,1}=-\boldsymbol{R}T^2B'(CO), \qquad \ldots\ldots(30)$$

$$E_{CO,2}=-\tfrac{1}{2}\boldsymbol{R}T^2C'(CO). \qquad \ldots\ldots(31)$$

Similarly,

$$H(T,V)=\Sigma[CO]\left\{H_{CO,0}(T)+\left(\frac{n}{V}\right)H_{CO,1}(T)+\left(\frac{n}{V}\right)^2 H_{CO,2}(T)\right\} \qquad \ldots\ldots(32)$$

with

$$H_{CO,1}=\boldsymbol{R}T\{B(CO)-TB'(CO)\}, \qquad \ldots\ldots(33)$$

$$H_{CO,2}=\tfrac{1}{2}\boldsymbol{R}T\{2C(CO)-TC'(CO)\}. \qquad \ldots\ldots(34)$$

Formulae for C_V and C_P can be derived from those in § 3.

In § 11 the quantities E_1, E_2, H_1 and H_2 are tabulated for the main products of combustion at high pressures.

The other two well-known combination rules are the "square-root" and Lorentz rules (cf. Beattie and Stockmayer, 1940). These involve squares and products of the mole numbers such as $[CO_2]$, and greatly increase the labour in the evaluation of pressures, internal energies, and heat contents. It is doubtful, too, whether the real accuracy of our calculations for the individual gases is sufficient to justify a superstructure of elaborate combination rules. To examine the order of magnitude of the changes involved, we have calculated the second virial coefficient for the products of a typical cordite. The square-root rule gave a value smaller by 8·5%, the Lorentz rule a value smaller by 5·5%. Probably the latter is the more accurate. The latter change corresponds to a 2% change in pressure at 30 tons/sq. in. The sign of the change indicates that the agreement with closed-vessel experiments would be made worse by the use of the more complicated rules.

The third virial coefficient has such a small effect on our thermochemical calculations that there is no need to look for a combination rule more accurate than the linear sum. This is also quite plausible.

§ 9. THE EQUILIBRIUM CONSTANT FOR THE WATER-GAS REACTION

From (27), (28) and (7) the Helmholtz free energy of 1 gm. of the mixture is

$$A=\Sigma[\mathrm{CO}]A_{\mathrm{CO}}(T)+\Sigma[\mathrm{CO}]\boldsymbol{R}T\ln\{[\mathrm{CO}]/V\}+\frac{n\boldsymbol{R}T}{V}\Sigma[\mathrm{CO}]B(\mathrm{CO})$$
$$+\left(\frac{n}{V}\right)^2\frac{\boldsymbol{R}T}{2}\Sigma[\mathrm{CO}]C(\mathrm{CO}). \qquad \ldots\ldots(35)$$

The chemical potential of CO in the mixture is

$$\mu_{\mathrm{CO}}=\left\{\frac{\partial A}{\partial[\mathrm{CO}]}\right\}_{T,\,V,\,[\mathrm{CO}_2],\,\ldots\ldots}$$
$$=A_{\mathrm{CO}}(T)+\boldsymbol{R}T\ln\{[\mathrm{CO}]/V\}+\boldsymbol{R}T+\frac{n\boldsymbol{R}T}{V}B(\mathrm{CO})+\frac{\boldsymbol{R}T}{V}\Sigma[\mathrm{CO}]B[\mathrm{CO}]$$
$$+\left(\frac{n}{V}\right)^2\frac{\boldsymbol{R}T}{2}C(\mathrm{CO})+\frac{n\boldsymbol{R}T}{V^2}\Sigma[\mathrm{CO}]C(\mathrm{CO}). \qquad \ldots\ldots(36)$$

The fugacity $P^*(\mathrm{CO})$ is therefore given by

$$\boldsymbol{R}T\ln P^*(\mathrm{CO})=\boldsymbol{R}T\ln\{[\mathrm{CO}]\boldsymbol{R}T/V\}+\frac{n\boldsymbol{R}T}{V}B(\mathrm{CO})+\frac{\boldsymbol{R}T}{V}\Sigma[\mathrm{CO}]B(\mathrm{CO})$$
$$+\left(\frac{n}{V}\right)^2\frac{\boldsymbol{R}T}{2}C(\mathrm{CO})+\frac{n\boldsymbol{R}T}{V^2}\Sigma[\mathrm{CO}]C(\mathrm{CO}),$$

and hence

$$P^*(\mathrm{CO})=\frac{\boldsymbol{R}T[\mathrm{CO}]}{V}\exp\left[\frac{nB(\mathrm{CO})}{V}+\left(\frac{n}{V}\right)^2\frac{C(\mathrm{CO})}{2}+\frac{1}{V}\Sigma[\mathrm{CO}]B(\mathrm{CO})\right.$$
$$\left.+\frac{n}{V^2}\Sigma[\mathrm{CO}]C(\mathrm{CO})\right]. \qquad \ldots\ldots(37)$$

The equilibrium of the water-gas reaction is determined by

$$\frac{P^*(\mathrm{CO})P^*(\mathrm{H_2O})}{P^*(\mathrm{CO_2})P^*(\mathrm{H_2})}=K_0(T), \qquad \ldots\ldots(38)$$

where K_0 is the equilibrium constant at low densities.

Hence

$$\frac{[\mathrm{CO}][\mathrm{H_2O}]}{[\mathrm{CO_2}][\mathrm{H_2}]}=K_0(T)\exp\left\{-\frac{n\Delta B}{V}-\frac{n^2\Delta C}{2V^2}\right\}, \qquad \ldots\ldots(39)$$

where

$$\Delta B=B(\mathrm{CO})+B(\mathrm{H_2O})-B(\mathrm{CO_2})-B(\mathrm{H_2}) \qquad \ldots\ldots(40)$$

and

$$\Delta C=C(\mathrm{CO})+C(\mathrm{H_2O})-C(\mathrm{CO_2})-C(\mathrm{H_2}). \qquad \ldots\ldots(41)$$

These are negative and of the same order of magnitude as the virial coefficients of carbon monoxide. The deviations from the perfect gas law therefore increase the effective equilibrium constant by as much as 50% at the highest densities attained in ballistics. ΔB and ΔC are tabulated in § 11.

§ 10. MINOR PRODUCTS

The ratios of the major products are determined by the equilibrium constant of the water-gas reaction and the composition of the fuel. The minor products [OH], [H], and [NO] are determined by the following equations

$$\frac{[\mathrm{OH}]\sqrt{[\mathrm{H_2}]}}{[\mathrm{H_2O}]}=\sqrt{\frac{V}{RT}}K_1(T)\exp\left[\frac{n}{V}\{B(\mathrm{H_2O})-B(\mathrm{OH})-\tfrac{1}{2}B(\mathrm{H_2})\}\right], \ldots\ldots(42)$$

$$\frac{[\mathrm{NO}][\mathrm{H_2}]}{[\mathrm{H_2O}]\sqrt{[\mathrm{N_2}]}}=\sqrt{\frac{V}{RT}}K_2(T)\exp\left[\frac{n}{V}\{B(\mathrm{H_2O})+\tfrac{1}{2}B(\mathrm{N_2})-B(\mathrm{NO})-B(\mathrm{H_2})\}\right], \ldots\ldots(43)$$

$$[\mathrm{H}]/\sqrt{[\mathrm{H_2}]}=\sqrt{\frac{V}{RT}}K_3(T)\exp\left[\frac{n}{V}\{\tfrac{1}{2}B(\mathrm{H_2})-B(\mathrm{H})\}\right]. \qquad \ldots\ldots(44)$$

K_1, K_2 and K_3 are functions listed in the literature.

The terms involving third virial coefficients have been neglected. We need B for OH, H and NO. We must keep in mind the fact that the pressure-corrections in (42), (43), and (44) will have only a small influence on thermochemical calculations, and therefore some attention may be paid to convenience in computation. As reasonable rough values for the second virial coefficients at high temperatures we take $B(\mathrm{H})\simeq 10$ c.c./mole ; $B(\mathrm{OH})\simeq 20$ c.c./mole; $B(\mathrm{NO})\simeq 30$ c.c./mole and then

$$B(\mathrm{H_2O})-B(\mathrm{OH})-\tfrac{1}{2}B(\mathrm{H_2})\simeq -20 \text{ c.c./mole},$$

$$B(\mathrm{H_2O})+\tfrac{1}{2}B(\mathrm{N_2})-B(\mathrm{NO})-B(\mathrm{H_2})\simeq -20 \text{ c.c./mole},$$

while $\tfrac{1}{2}B(\mathrm{H_2})-B(\mathrm{H})$ is only a few c.c./mole over the temperature ranges at which these species are present. Hence we can write, with sufficient accuracy,

$$\frac{[\mathrm{OH}]\sqrt{[\mathrm{H_2}]}}{[\mathrm{H_2O}]}=\sqrt{\frac{V}{RT}}K_1(T)\exp(-20\,n/V), \qquad \ldots\ldots(45)$$

$$\frac{[\mathrm{NO}][\mathrm{H_2}]}{[\mathrm{H_2O}]\sqrt{[N_2]}}=\sqrt{\frac{V}{RT}}K_2(T)\exp(-20\,n/V), \qquad \ldots\ldots(46)$$

$$[\mathrm{H}]/\sqrt{[\mathrm{H_2}]}=\sqrt{\frac{V}{RT}}K_3(T). \qquad \ldots\ldots(47)$$

It is convenient that two of the pressure-corrections are the same and the third is unity.

Of the other products of propellant explosions, O_2, O, and N occur in such small quantities that the relevant equilibrium constants need not be corrected for gas-imperfection. Methane and ammonia also occur in only small amounts. There would be no difficulty in calculating high-temperature virial coefficients for these gases; methane is effectively spherical and ammonia is covered by Stockmayer's work (Stockmayer, 1941).

In practical calculations based on our tables we have neglected the contribution of the minor products to the virial coefficients of the mixture. Taking the rough values used above, we have found that the inclusion of these terms would change the pressure by only 0·2% in the worst case. Similarly the minor products make only a small contribution to the pressure-dependent parts of the internal energy and heat content of the product mixture.

§ 11. TABLES OF RESULTS

The notation has been explained in §§ 3 and 9. The number of places kept in these tables is greater than the real accuracy, and was chosen to give products of combustion smooth to $\frac{1}{2} \times 10^{-5}$ mole/gm. and temperatures smooth to one or two degrees.

(*a*) Tables for pressures

Temp. (°K)	Table of *B* (c.c./mole)				Table of *C* (c.c./mole)^2			
	H_2	N_2, CO	CO_2	H_2O	H_2	N_2, CO	CO_2	H_2O
1600	16·4	32·1	45·7	−4·2	20	210	1385	220
1700	16·3	32·3	47·3	−2·5	20	200	1305	210
1800	16·2	32·4	48·7	−1·1	20	190	1235	195
1900	16·1	32·6	49·9	+0·2	20	180	1170	185
2000	16·0	32·6	50·9	1·2	15	170	1110	175
2100	15·9	32·7	51·8	2·2	15	160	1055	170
2200	15·8	32·7	52·6	3·0	15	155	1010	160
2300	15·7	32·8	53·2	3·7	15	150	965	155
2400	15·6	32·8	53·8	4·4	15	140	925	145
2500	15·6	32·8	54·4	5·0	15	135	885	140
2600	15·5	32·7	54·8	5·5	15	130	855	135
2700	15·4	32·7	55·3	6·0	15	125	825	130
2800	15·3	32·7	55·6	6·4	10	120	795	125
2900	15·3	32·6	56·0	6·8	10	120	765	120
3000	15·2	32·6	56·2	7·1	10	115	740	120
3100	15·1	32·6	56·5	7·5	10	110	720	115
3200	15·0	32·5	56·7	7·7	10	105	695	110
3300	15·0	32·4	56·9	8·0	10	105	675	105
3400	14·9	32·4	57·1	8·3	10	100	650	105
3500	14·8	32·3	57·3	8·5	10	95	635	100
3600	14·8	32·3	57·4	8·7	10	95	615	100
3700	14·7	32·2	57·5	8·9	10	90	600	95
3800	14·7	32·2	57·6	9·1	10	90	585	95
3900	14·6	32·1	57·7	9·3	10	85	570	90
4000	14·5	32·0	57·8	9·4	10	85	555	90

Reference to the tables of E_1 will show that the values for the different gases include both signs and accordingly the summation over all constituents tends to be small. E itself is of order 600–1000 cal./gm., and of this about -3 cal./gm. comes from the E_1 terms and about $+1$ cal./gm. from the E_2 terms (both these figures refer to pressures of order 20 tons/sq. in.). In the calculations of internal ballistics H is of order 600–1000 cal./gm., while H_1 terms contribute about 40 cal./gm. and the H_2 terms 4 cal./gm. at 20 tons/sq. in. This demonstrates how small are the pressure-corrections to the internal energy, compared with those to the heat content. As a corollary, calculations without pressure-corrections give temperatures of uncooled explosion correct to within a few degrees; on the other hand, the temperature of burning at constant pressure is really lower than that computed in such a simple manner, to the extent of 150° c. at 20 tons/sq. in.

(*b*) Tables for equilibrium constants
Definitions in equations (39)–(41)

Temperature (°K.)	$-\Delta B$ (c.c./mole)	$-\frac{1}{2}\Delta C$ (c.c./mole)^2
1600	34·2	490
1700	33·8	460
1800	33·5	435
1900	33·2	410
2000	33·0	390
2100	32·8	370
2200	32·6	355
2300	32·5	340
2400	32·4	325
2500	32·2	310
2600	32·1	300
2700	32·0	290
2800	31·9	280
2900	31·8	270
3000	31·7	260
3100	31·6	255
3200	31·5	245
3300	31·4	235
3400	31·4	230
3500	31·3	225
3600	31·2	215
3700	31·1	210
3800	31·1	205
3900	31·0	200
4000	30·9	195

C_{V_2} has maximum values of less than 40 (cal./deg.)(c.c./mole)^2 for H_2O, 20 for CO_2, 1 for N_2 and CO, and 0·1 for H_2. It has therefore not been tabulated. β, defined in (13), is less than 6 (c.c./mole)^2 for H_2O, 4 for CO_2, 2×10^{-4} for N_2 and CO, and 5×10^{-5} for H_2.

(*c*) Tables for internal energies and heats
Definitions in equations (3)–(6)

Temp. (°K.)	$E_1/100$ (cal.–c.c./mole)				$H_1/100$ (cal.–c.c./mole)			
	H_2	N_2, CO	CO_2	H_2O	H_2	N_2, CO	CO_2	H_2O
1600	45	−110	−870	−935	565	910	585	−1065
1700	50	−95	−840	−895	600	995	760	−975
1800	55	−80	−810	−855	635	1080	930	−895
1900	65	−70	−785	−825	670	1160	1100	−820
2000	70	−55	−755	−795	705	1240	1265	−745
2100	75	−40	−730	−770	740	1325	1430	−680
2200	80	−25	−700	−745	775	1405	1595	−615
2300	90	−15	−675	−725	805	1485	1760	−555
2400	95	0	−645	−705	840	1560	1920	−495
2500	100	15	−620	−685	875	1640	2080	−440
2600	105	25	−595	−665	905	1720	2240	−385
2700	110	40	−565	−650	940	1795	2400	−330
2800	120	55	−540	−635	970	1870	2555	−280
2900	125	65	−515	−620	1005	1950	2710	−230
3000	130	80	−490	−605	1035	2025	2865	−180
3100	135	95	−465	−590	1065	2100	3015	−135
3200	140	105	−440	−580	1100	2170	3170	−85
3300	145	120	−415	−565	1130	2245	3320	−40
3400	150	130	−390	−555	1160	2320	3470	+5
3500	160	145	−365	−540	1190	2395	3620	50
3600	165	155	−340	−530	1220	2465	3770	90
3700	170	170	−315	−520	1250	2540	3920	135
3800	175	185	−290	−510	1280	2610	4065	175
3900	180	195	−265	−500	1310	2680	4210	220
4000	185	210	−240	−490	1340	2750	4355	260

At all temperatures in this range use

	H_2	N_2, CO	CO_2	H_2O
$10^{-4}E_2$ (cal. (c.c./mole)^2)	3	34	220	35
$10^{-4}H_2$ (cal. (c.c./mole)^2)	10	101	661	106

(*d*) Tables for specific heats

Symbols defined in equations (8) – (13).

Temp. (°K.)	C_{V_1} (cal.–c.c./mole-deg.)				a (c.c./mole)			
	H_2	N_2, CO	CO_2	H_2O	H_2	N_2, CO	CO_2	H_2O
1600	6·4	14·3	29·4	43·8	15·0	35·6	73·1	25·2
1700	6·3	14·1	28·9	38·2	14·8	35·1	72·2	23·9
1800	6·3	14·0	28·4	33·9	14·6	34·7	71·4	22·9
1900	6·2	13·9	28·0	30·3	14·4	34·4	70·6	22·0
2000	6·2	13·8	27·7	27·4	14·2	34·0	69·9	21·3
2100	6·1	13·7	27·4	24·9	14·1	33·7	69·2	20·6
2200	6·1	13·7	27·1	22·9	14·0	33·4	68·6	20·1
2300	6·1	13·6	26·9	21·1	13·8	33·0	68·0	19·6
2400	6·0	13·5	26·7	19·6	13·7	32·7	67·4	19·1
2500	6·0	13·4	26·5	18·3	13·6	32·5	66·9	18·7
2600	5·9	13·4	26·3	17·2	13·5	32·2	66·3	18·4
2700	5·9	13·3	26·2	16·2	13·3	32·0	65·8	18·1
2800	5·9	13·2	26·0	15·4	13·2	31·7	65·4	17·8
2900	5·8	13·1	25·8	14·6	13·1	31·5	64·9	17·5
3000	5·8	13·1	25·7	13·9	13·0	31·3	64·5	17·3
3100	5·8	13·0	25·6	13·3	12·9	31·0	64·0	17·1
3200	5·7	13·0	25·5	12·8	12·8	30·8	63·6	16·9
3300	5·7	12·9	25·4	12·3	12·7	30·6	63·2	16·7
3400	5·7	12·9	25·3	11·8	12·7	30·4	62·9	16·5
3500	5·6	12·8	25·1	11·4	12·6	30·3	62·5	16·3
3600	5·6	12·8	25·0	11·0	12·5	30·1	62·1	16·1
3700	5·6	12·7	24·9	10·7	12·4	29·9	61·8	16·0
3800	5·6	12·7	24·9	10·4	12·4	29·7	61·5	15·9
3900	5·5	12·6	24·8	10·1	12·3	29·6	61·1	15·7
4000	5·5	12·6	24·7	9·9	12·2	29·4	60·8	15·6

§ 12. COMPARISON WITH EXPERIMENT

We have made a detailed study of the products of explosion of cordite " X ".

(a) *Explosion of " X " without performance of external work*

The following numerical data were taken as *exact*: atomic composition {C} = 2238; {H} = 3015; {N} = 1044; {O} = 3468, all in 10^{-5} gm.-atom/gm.; heat of formation of propellant at constant volume = 470 cal./gm. From these data we have calculated the products of explosion, first from Dr. Pike's tables alone, and then with pressure-corrections from the tables given earlier in this paper. From the pressure of the products, we derive the " covolume " of the propellant by the definition

$$P(V-\eta)=n\boldsymbol{R}T_0, \qquad \text{......(48)}$$

where T_0 and n have been calculated at 4000 atmospheres from tables without pressure-corrections.

Table 8 shows the theoretical temperature, products, pressure and covolume at various product densities. In figures 2 and 3 are shown the behaviour of T_0, and the variation of one of the main products. T_0 differs by no more than 3° C. from that calculated without pressure-corrections. The main products are considerably different. The most interesting result in table 8 is the " theoretical covolume" η. At pressures over 5 tons/sq. in., η decreases with increase of density. The accepted value for this cordite is 0·94 c.c./gm., determined by closed-vessel firings at 27 tons/sq. in. Our value at this pressure is 5% too small.

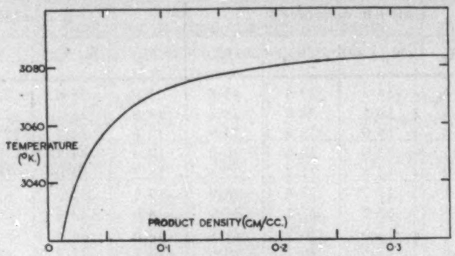


Figure 2. Product temperature in uncooled explosion of "X" without performance of external work.

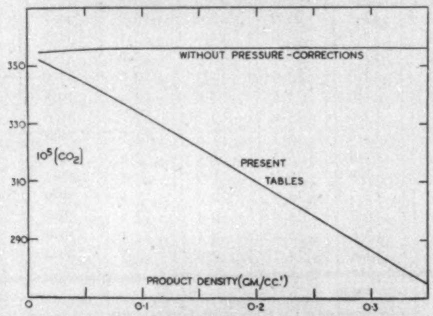


Figure 3. Variation of a major product with density in explosion of "X" without performance of external work.

As the maximum pressure is reduced below 5 tons/sq. in. the covolume begins to decrease again, until it becomes negative at pressures of order 1 ton/sq. in. This behaviour is due to the increasing dissociation causing low temperatures of explosion.

From our computations for the products of "X" at temperatures from 2000 to 3100° K., it can be shown that

$$P = \frac{nRT}{V-b} \qquad \ldots\ldots(49)$$

is, at each temperature, a less accurate representation than

$$P = \frac{nRT}{V}\left(1 + \frac{\beta}{V}\right). \qquad \ldots\ldots(50)$$

This had been predicted before the present tables were started, as it follows immediately from an inspection of the ratio (third virial coefficient)/(second virial coefficient)^2 at high temperatures. Using the best values of b and β, the

Table 8. Explosion of " X " at constant volume

Product density (gm./c.c.)	0·01	0·05	0·1	0·2	0·3	0·35
T_0, (°K.)	3018	3058	3073	3081	3081	3083
[CO_2] in 10^{-5} mole/gm.	352	343	333	310	286	275
[CO]	1886	1895	1905	1928	1952	1963
[H_2O]	853	873	888	914	938	951
[H_2]	628½	622	610	587	563	552
[N_2]	521½	521½	522	522	522	522
[OH]	23	12	8½	5	5	4
[H]	27	14	10	7	6	5
[NO]	1	1	½	½	0	0
$10^5 n$	4292	4282	4277	4274	4272	4272
P (atmos.)	107·5	568·2	1204	2672	4407	5381
P (tons/sq. in.)	0·705	3·73	7·90	17·53	28·91	35·30
Covolume, c.c./gm.	−0·5	0·98	1·02	0·96	0·88	0·85

errors in (49) and (50) rise to the order of 2% and 0·7% respectively. Both b and β vary little with T, certainly by less than 1% over the range 2000–3000° K.

Our tables have been applied with success to the calculation of the behaviour of other propellants, and show a systematic error of about 4% in the covolume.

(b) *Explosion of " X " under constant pressure*

The results of computations from our tables are shown in table 9 and figures 4 and 5. The temperature of the products, T_m, has a maximum just below 1 ton/sq. in., decreasing at smaller pressures because of the increasing dissociation

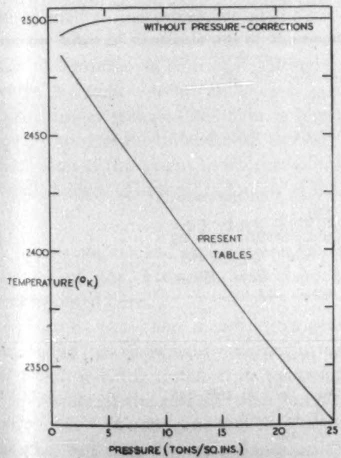


Figure 4. Theoretical product temperatures for explosion of " X " under constant-pressure conditions.

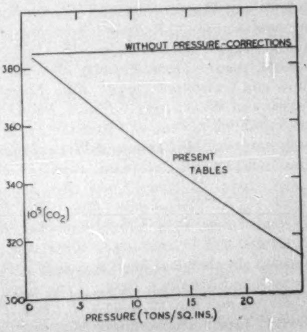


Figure 5. Composition of products of " X " under constant-pressure conditions.

and decreasing at higher pressures because of the increasing heat content of the imperfect gases. At 25 tons/sq. in. our T_m is about 180° c. lower than that calculated without pressure-corrections.

Table 9. Explosion of " X " under constant pressure

Pressure (atmos.)	88·2	457·8	956·6	2074·7	3351	4048
,, (tons/sq. in.)	0·58	3·00	6·28	13·61	21·99	26·56
Density of products (gm./c.c.)	0·01	0·05	0·1	0·2	0·3	0·35
T_m (°K.)	2488	2472	2448	2396	2343	2317
$[CO_2]$ in 10^{-5} mole/gm.	385	377	366	344	321	309
$[CO]$	1853	1861	1872	1894	1917	1929
$[H_2O]$	843	852	864	886	909	921
$[H_2]$	662	654	643	621	598	587
$[N_2]$	522	522	522	522	522	522
$[OH]$	2	1	0	0	0	0
$[H]$	4	2	1	1	1	0
$10^5 n$	4271	4269	4268	4268	4268	4268

ACKNOWLEDGMENTS

I am indebted to Dr. H. H. M. Pike for placing at my disposal his knowledge of both experimental and theoretical aspects of closed-vessel technique, and to the Chief Scientific Officer, Ministry of Supply, for permission to publish this paper.

REFERENCES

ADEL and DENISON, 1935. *Phys. Rev.*, **48,** 516.
AMAGAT, 1893. *Ann. chim. phys.*, **29,** 109.
BEATTIE and BRIDGEMAN, 1928. *J. Amer. Chem. Soc.*, **50,** 3151.
BEATTIE and STOCKMAYER, 1940. *Rep. Progr. Phys.*, **7,** 195.
BIRGE, 1941. *Rev. Mod. Phys.*, **13,** 233.
DE BOER and MICHELS, 1938. *Physica*, **5,** 945 ; 1939. *Ibid.*, **6,** 97.
BRIDGEMAN, 1927. *J. Amer. Chem. Soc.*, **49,** 1130.
BURLOT, 1932. *Mem. Poudres*, **25,** 314.
CORNER, 1940. *Trans. Faraday Soc.*, **36,** 780 ; 1946. In the press.
CROW and GRIMSHAW, 1931. *Phil. Trans.*, A, **230,** 39.
GRIMM and WOLF, 1933. *Handb. Physik*, **24** (2nd edition), pt. 1, 977.
HIRSCHFELDER, EWELL and ROEBUCK, 1938. *J. Chem. Phys.*, **6,** 205.
HIRSCHFELDER, MCCLURE and WEEKS, 1942. *J. Chem. Phys.*, **10,** 201.
HIRSCHFELDER and ROSEVEARE, 1939. *J. Phys. Chem.*, **43,** 15.
KASSEL, 1934. *J. Amer. Chem. Soc.*, **56,** 1838.
KEYES, 1941. *Temperature* (Amer. Inst. Phys., N.Y.), p. 45.
LENNARD-JONES, 1924. *Proc. Roy. Soc.*, A, **106,** 463 ; 1931. *Proc. Phys. Soc.*, **43,** 461.
MARGENAU and MYERS, 1944. *Phys. Rev.*, **66,** 307.
MICHELS, BLAISSE and MICHELS, 1937. *Proc. Roy. Soc.*, A, **160,** 358.
MICHELS and MICHELS, 1935. *Proc. Roy. Soc.*, **A, 153,** 201.
MONTROLL and MAYER, 1941. *J. Chem. Phys.*, **9,** 626.
NEWITT, 1940. *High-pressure plant and fluids at high pressures*, p. 191.
ROEBUCK, MURRELL and MILLER, 1942. *J. Amer. Chem. Soc.*, **64,** 400.
STOCKMAYER, 1941. *J. Chem. Phys.*, **9,** 398.
WHITELAW, 1934. *Physica*, **1,** 749.

ON THE ZERNIKE PHASE-CONTRAST TEST

By E. H. LINFOOT,
University of Bristol

Communicated by E. W. H. Selwyn ; *MS. received* 1 *June* 1946

ABSTRACT. A diffraction theory of the Zernike test is developed for the case where the dimensions of the light source are negligibly small, and applied to estimate the sensitiveness of the test for detecting local zones in an otherwise true mirror. The effects of varying the size of the phase-retarding disc on the appearance under test of a true mirror are considered, and the conclusion is drawn that it is these effects, rather than considerations of sensitivity, which mainly determine the upper limit to the useful size of the disc.

---

§ 1. INTRODUCTION

An account of the diffraction theory of the phase-contrast method of testing the figures of optical surfaces has been given in two interesting papers by its originator, F. Zernike (1934 a, b). However, the mathematical analysis in these papers takes a form which is not very well adapted to a quantitative discussion of the properties of the test. Moreover, it is for the most part restricted to surfaces whose errors are small compared with $\lambda/4\pi$, where λ is the wave-length of the light, that is to say, to surfaces which are already practically perfect.

In the present paper we re-develop the theory of the phase-contrast test and obtain formulae which predict the appearance under test of a mirror with arbitrary smooth errors of figure, not necessarily small compared with λ. These formulae are applied to the discussion of the appearance of a true mirror under the test and the limit of sensitiveness of the test for local zonal errors. The analysis, like that of Zernike, is restricted throughout to the mathematically simplest case where the light-source (pinhole) is small compared with the Airy disc of the mirror under test and the light is monochromatic. In practice, the pinhole is usually comparable in size with the Airy disc and the light polychromatic, but a discussion of the extent to which the properties of the test are modified by these circumstances is beyond the scope of the present paper.

§ 2. THE PHASE-CONTRAST TEST

The notation and assumptions are similar to those adopted in a recent discussion of the Foucault test (Linfoot, 1946). For convenience they are recapitulated here.

All the cases where a convergent pencil is being null-tested can be covered by supposing that the wave-fronts originate at the surface of a nearly spherical mirror M which is being tested at its approximate centre of curvature. See figure 1. The results of the test are then interpreted in terms of errors of figure of the mirror M. We suppose that these errors of figure may amount to several wavelengths, but that the error-slopes on the mirror, besides being free from discontinuities, are not so steep as to spread out the visible image to more than a moderate multiple, say 5 or 10, of the size of the Airy disc. Then it makes no appreciable difference if we suppose the wave to be leaving a true spherical surface M_0, lying everywhere within a few wave-lengths of the surface M, and suppose that at the point (x, y, z) on M_0 the complex displacement is

$$E(x, y) = |E(x, y)|\, e^{\frac{2\pi i}{\lambda}\phi(x, y)};$$

thus $|E(x, y)|$ is the amplitude at the point (x, y) on M_0 and $\phi(x, y)$ the phase there; the variation in $\phi(x, y)$ expresses the distortion of the wave-fronts. Outside the boundary-circle C, lying in M_0 and corresponding to the edge of the mirror, we define $E(x, y)$ as zero.

After leaving the mirror, the light comes to a more or less imperfect focus in the neighbourhood of the centre of curvature O of M_0, at a distance s from its pole A; we suppose that s is large compared with the diameter of the mirror. Just behind O, a viewing system L receives the light and forms an image M_0' of the surface M_0; we denote by C′ the boundary of M_0'.

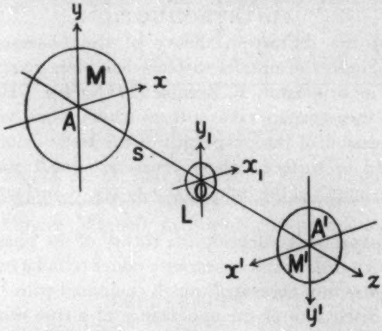


Figure 1.

At O is placed the phase-retarding Zernike disc, comparable in size with the Airy disc of the mirror, which renders variations of phase in the wave-fronts leaving M_0 visible as variations of intensity in the image-surface M_0'. x_1 and y_1, measured parallel to x and y, are coordinates in the space near O.

We define a system of coordinate-numbers (x', y') in the surface M_0' by assigning to the image P′ of the point $P = (x, y, z)$ in M_0 the coordinate-numbers $x' = x$, $y' = y$. The intensity at P′ is thus the same thing as the " intensity seen under the test " at P. To avoid unessential complications, we suppose that L is a perfect optical system.

By the usual arguments, based on an application of Huyghens' principle,[*] the complex displacement in the intermediate image-surface, which we take to be the sphere S of centre A passing through O, is

$$W(u, v) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ixu+iyv} E(x, y)\,dx\,dy, \qquad \ldots\ldots(2.1)$$

where

$$u = \frac{2\pi x_1}{\lambda s}, \quad v = \frac{2\pi y_1}{\lambda s}, \qquad \ldots\ldots(2.2)$$

[*] See the paper already referred to for a fuller account of these arguments.

and the integral is only formally over an infinite domain, since the integrand vanishes outside the boundary-contour C.

If the phase-retarding disc were not present, the complex displacement $D(x', y')$ at the point (x', y') in the final image-surface would be obtained by integrating the product

$$\frac{1}{2\pi} e^{-i(x'u+y'v)} W(u, v)$$

over the area $u^2+v^2 \leqq R^2$ of the (u, v)-plane which falls within the aperture of the viewing system L.* The effect of a disc occupying the region $u^2+v^2 \leqq a^2$ of the intermediate image-surface and retarding by α radians the phase of the waves passing through it can be represented by changing $W(u, v)$ into $e^{-i\alpha}W(u, v)$ throughout this region and leaving it unaltered elsewhere. In saying this, we assume that the effect of the disc is merely to change the phase of those parts of the wave which impinge upon it, without influencing the neighbouring parts. This assumption is never strictly true, but it is a permissible approximation when the pencil under test is of small numerical aperture.†

The complex displacement at the point (x', y') of the final image-surface is then given, to a sufficient approximation, by the equation

$$D(x', y') = \frac{1}{2\pi} \iint\limits_{u^2+v^2 \leqq R^2} e^{-iux'-ivy'} W(u, v)\, du\, dv + \frac{1}{2\pi}(e^{-i\alpha}-1) \iint\limits_{u^2+v^2 \leqq a^2} e^{-iux'-ivy'} W(u, v)\, du\, dv. \qquad \ldots\ldots(2.3)$$

The first term of (2.3) is simply the complex displacement which results when the mirror is imaged through the viewing system, without the interposition of the Zernike disc, and the second term therefore describes the effect of the disc.

We next replace R by ∞ in the first term of (2.3). This amounts to neglecting the effects of the finite aperture of the viewing telescope. For mirrors with errors of the type under consideration, the resulting change in the predicted intensities is too small to be detected by the eye, except at points whose distance from the rim of the mirror is comparable with the resolving power of the viewing system. At these points the intensities depend critically on the aperture and optical aberrations of the viewing system, as well as on the errors of the mirror, and we therefore exclude them from the discussion. At the remaining points we obtain, on setting $R=\infty$ in (2.3), the equation

$$\begin{aligned} D(x', y') &= \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-iux'-ivy'} W(u, v)\, du\, dv \\ &\qquad + \frac{1}{2\pi}(e^{-i\alpha}-1) \iint\limits_{u^2+v^2 \leqq a^2} e^{-iux'-ivy'} W(u, v)\, du\, dv \\ &= E(x', y') + \frac{1}{2\pi}(e^{-i\alpha}-1) \iint\limits_{u^2+v^2 \leqq a^2} e^{-iux'-ivy'} W(u, v)\, du\, dv \qquad \ldots\ldots(2.4) \end{aligned}$$

---

* In the case of a $f/10$ pencil and a viewing telescope of 1-inch aperture the value of R is approximately 1300.

† For in this case the scale of the diffraction pattern in S is large compared with the wave-length λ. The radius of the first Airy dark ring is 3·83 in (u, v)-units.

for the complex displacement $D(x', y')$. The intensity seen under the test, at the point (x', y') on the mirror-surface or in the halo, is then measured by the quantity

$$I(x', y') = |D(x', y')|^2. \qquad \ldots\ldots(2.5)$$

§ 3. TRANSFORMATION OF (2.4): GENERAL FORMULAE

The second term in (2.4) describes the change in $D(x', y')$ caused by the presence of the Zernike disc. On substituting for $W(u, v)$ from (2.1) and remembering that $E(x, y)$ vanishes outside the part of M_0 corresponding to the mirror-surface, we can write this term in the form

$$\frac{1}{4\pi}(e^{-i\alpha}-1)\iint\limits_{u^2+v^2\leqq a^2} e^{-iux'-ivy'}\,du\,dv \iint\limits_{\mathrm{M}} e^{ixu+iyv}E(x, y)\,dx\,dy$$

$$= \frac{1}{4\pi}(e^{-i\alpha}-1)\iint\limits_{\mathrm{M}} E(x, y)\,dx\,dy \iint\limits_{u^2+v^2\leqq a^2} e^{iu(x-x')+iv(y-y')}\,du\,dv. \qquad \ldots\ldots(3.1)$$

The inner integral

$$\iint\limits_{u^2+v^2\leqq a^2} e^{iu(x-x')+iv(y-y')}\,du\,dv = 2\pi a^2\frac{J_1(aZ)}{aZ},$$

where $Z = \sqrt{[(x-x')^2+(y-y')^2]}$ and J_1 is the Bessel function of order 1. Hence, (2.4) can be written in the form

$$D(x', y') = E(x', y') + \frac{a^2}{2\pi}(e^{-i\alpha}-1)\iint\limits_{\mathrm{M}} E(x, y)\frac{J_1(aZ)}{aZ}\,dx\,dy. \qquad \ldots\ldots(3.2)$$

It is possible to obtain (3.2) by arguments of a more physical character. The Zernike disc can be regarded as sending out a "virtual" supplementary wave (namely, one produced backwards in time) which interferes at the surface M_0 with the original wave $E(x', y')$. The resulting complex displacement $D(x', y')$ on M_0 determines the observed intensity $I(x', y') = |D(x', y')|^2$. Many of the properties of the Zernike test can be derived qualitatively by regarding the supplementary wave as a spherical wave originating at the centre of the Zernike disc,* but this simplified picture does not, of course, suffice for the derivation of (3.2). The idea can be carried through quantitatively, however, by applying the principle of superposition in the following way. Each element $dx\,dy$ of the mirror-surface sends out a spherical wave which is partly intercepted by the Zernike disc. The complex displacement which this wave causes in the surface of a sphere centred at the element $dx\,dy$ and passing through O (see figure 1) is $\frac{i}{\lambda s}e^{-\frac{2\pi i}{\lambda}s}E(x, y)\,dx\,dy$. The corresponding contribution to the virtual wave emitted by the Zernike disc is represented, to a sufficient approximation, by a complex displacement $\frac{i}{\lambda s}e^{-\frac{2\pi i}{\lambda}s}(e^{-i\alpha}-1)E(x, y)\,dx\,dy$ over the region of this sphere which most nearly coincides with the area of the disc, namely the circle of arcual centre O and of radius $\frac{\lambda s}{2\pi}a$. Now, by the classical theory of the Airy disc, a real wave-front filling this circle and with complex displacement-function equal to 1 over its area would produce at the point (x', y') on M_0 a complex displacement

---

* See C. R. Burch (1934).

$$\frac{2\pi i}{\lambda s} e^{-\frac{2\pi i}{\lambda}s} \left(\frac{\lambda s}{2\pi} a\right)^2 \frac{J_1(aZ)}{aZ},$$

where $Z = \sqrt{[(x-x')^2 + (y-y')^2]}$. In the time-reversed "virtual" wave, the factor $\frac{2\pi i}{\lambda s} e^{-\frac{2\pi i}{\lambda}s}$ becomes $-\frac{2\pi i}{\lambda s} e^{\frac{2\pi i}{\lambda}s}$, while the initial complex displacement-function is not 1 but $\frac{i}{\lambda s} e^{-\frac{2\pi i}{\lambda}s}(e^{-i\alpha}-1)E(x,y)dx\,dy$. If follows that the contribution from the element $dx\,dy$ to the virtual supplementary wave sent out by the Zernike disc is such as to produce at a point (x', y') in the surface M_0 the complex displacement

$$\frac{i}{\lambda s} e^{-\frac{2\pi i}{\lambda}s}(e^{-i\alpha}-1)E(x,y)dx\,dy \,.\, -\frac{2\pi i}{\lambda s} e^{\frac{2\pi i}{\lambda}s} \left(\frac{\lambda s}{2\pi} a\right)^2 \frac{J_1(aZ)}{aZ}$$
$$= \frac{a^2}{2\pi}(e^{-i\alpha}-1)E(x,y)\frac{J_1(aZ)}{aZ}\,dx\,dy.$$

Integrating over the elements $dx\,dy$, we obtain

$$\frac{a^2}{2\pi}(e^{-i\alpha}-1)\iint\limits_{M} E(x,y)\frac{J_1(aZ)}{aZ}\,dx\,dy$$

as the total complex displacement at (x', y') due to the virtual wave from the Zernike disc, and (3.2) follows at once.

Evidently this argument can be applied to mirrors of arbitrary edge-contour and variable reflecting power; by taking α complex in the factor $(e^{-i\alpha}-1)$ we can also cover the case where the Zernike disc absorbs a known fraction of the light which passes through it. But since practical interest centres mainly on the circular mirror of uniform reflecting power, tested with a disc whose absorption and back-reflection can be disregarded, the discussion is here restricted to this case.

To calculate the intensities seen under the test on a mirror whose errors of figure correspond to the presence in the image of small amounts of the classical aberrations, we expand the function $E(x, y)$ as a series

$$E(x,y) = E(r\cos\phi, r\sin\phi) = \sum_{m=-\infty}^{\infty} E_m(r)e^{im\phi}$$

over the mirror-surface $0 \leqq r \leqq 1$, $0 \leqq \phi \leqq 2\pi$. The integral (3.1) can then be written

$$\frac{1}{4\pi^2}(e^{-i\alpha}-1)\int_0^1 r\,dr\int_0^a \rho\,d\rho\int_0^{2\pi} E(r\cos\phi, r\sin\phi)d\phi\int_0^{2\pi} e^{ir\rho\cos(\phi-\chi)-ir'\rho\cos(\phi'-\chi)}d\chi,$$

where $\qquad x+iy = re^{i\phi},\ x'+iy' = r'e^{i\phi'},\ u+iv = \rho e^{i\chi}$

$$= \frac{1}{2\pi}(e^{-i\alpha}-1)\int_0^a \rho\,d\rho\int_0^1 r\,dr\int_0^{2\pi} E(r\cos\phi, r\sin\phi)J_0(\rho Z)d\phi.$$

where
$$Z=\sqrt{[(x-x')^2+(y-y')^2]}$$
$$=\frac{1}{2\pi}(e^{-i\alpha}-1)\sum_m\int_0^a\rho\,d\rho\int_0^1 E_m(r)r\,dr\int_0^{2\pi}e^{im\phi}J_0(\rho Z)d\phi$$
$$=(e^{-i\alpha}-1)\sum_m\int_0^a\rho\,d\rho\int_0^1 E_m(r)J_{|m|}(\rho r)J_{|m|}(\rho r')r\,dr.$$

(3.2) thus becomes
$$D(x',y')=E(x',y')+(e^{-i\alpha}-1)\sum_{m=-\infty}^{\infty}e^{im\phi'}\int_0^1 E_m(r)r\,dr\int_0^a J_{|m|}(\rho r)J_{|m|}(\rho r')\rho\,d\rho$$
$$=E(x',y')+a^2(e^{-i\alpha}-1)\sum_{m=-\infty}^{\infty}e^{im\phi'}\int_0^1 E_m(r)K_{|m|}(ar,ar')r\,dr, \qquad \ldots\ldots(3.3)$$

where K_m is defined by the equations
$$K_m(u,v)=\frac{1}{u^2-v^2}[uJ_{m+1}(u)J_m(v)-vJ_{m+1}(v)J_m(u)] \qquad (u\neq v)$$
$$=-\frac{1}{2u}[uJ_{m+1}(u)J_m{}'(u)-uJ_m(u)J'_{m+1}(u)-J_m(u)J_{m+1}(u)] \qquad (u=v). \qquad \ldots\ldots(3.4)$$

More particularly, following Zernike, we may expand $E(x,y)$ as a series of polynomials in x and y, orthogonal over the unit circle, by writing
$$E(x,y)=\sum_{m,n}a_{m\,n}R_n{}^{|m|}(r)e^{im\phi}, \qquad \ldots\ldots(3.5)$$
where, in the summation, $-n\leqq m\leqq n$ and $m-n$ is even, while $R_n{}^m(r)$ is the polynomial
$$(-1)^{\frac{n-m}{2}}\binom{\frac{1}{2}(n+m)}{m}r^mF\left(\frac{n}{2}+\frac{m}{2}+1,-\frac{n}{2}+\frac{m}{2},m+1,r^2\right).$$

The equation for $D(x',y')$ then takes the form
$$D(x',y')=E(x',y')+(e^{-i\alpha}-1)\sum a_{mn}e^{im\phi'}\int_0^a J_{|m|}(\rho r')\rho\,d\rho\int_0^1 R_n^{|m|}(r)J_{|m|}(\rho r)r\,dr$$
$$=E(x',y')+(e^{-i\alpha}-1)\sum(-1)^{\frac{n-m}{2}}a_{mn}e^{im\phi'}\int_0^a J_{|m|}(\rho r')J_{n+1}(\rho)d\rho$$
$$=E(x',y')+(e^{-i\alpha}-1)\sum(-1)^{\frac{n-m}{2}}a_{mn}e^{im\phi'}I_{m,n}(r',a), \qquad \ldots\ldots(3.6)$$
where
$$I_{mn}(r',a)=\int_0^a J_{|m|}(\rho r')J_{n+1}(\rho)d\rho.$$

The true mirror

In the case of a true mirror tested with the Zernike disc at focus, $a_{00}=1$ and the remaining coefficients a_{mn} are zero; thus
$$D(x',y')=D(r')=\begin{matrix}1\\0\end{matrix}+(e^{-i\alpha}-1)I(r',a), \qquad \ldots\ldots(3.7)$$

where *

$$I(r', a) = I_{00}(r', a) = \int_0^a J_0(r'\rho)J_1(\rho)d\rho$$

$$= (1 - J_0(a))\frac{2J_1(ar')}{ar'} + (1 - J_0(a) - 2J_2(a))\frac{6J_3(ar')}{ar'}$$

$$+ (1 - J_0(a) - 2J_2(a) - 2J_4(a))\frac{10J_5(ar')}{ar'} + \dots, \qquad \dots\dots(3.8)$$

and in (3.7) the upper or lower alternative is to be taken according as $x'^2 + y'^2 \lessgtr 1$.

Figure 2 *a*, which is similar to one given by Zernike (1934 a), shows the intensities $I_0 = I_0(r') = |D(r')|^2$ seen on the mirror and in the halo with quarter-wave discs $\left(\alpha = \frac{\pi}{2}\right)$ of radii $a = 2{\cdot}5$, $a = 2{\cdot}0$ and $a = 1{\cdot}6$. It will be seen that in the first two cases the mirror appears brighter in the centre. In the practical use of the test, the focal setting of the Zernike disc is empirically chosen so as to make the intensities seen on the mirror surface as uniform as possible. It follows that in the first two cases the disc will be set, not at focus, but a little inside it. Figure 2 *b* † shows the intensities seen on a true mirror at the focal settings which would actually be used in practice for these three values of *a*; the amount of defocusing is about $\frac{1}{12}$ fringe in the first case and $\frac{1}{50}$ fringe in the second; in the third case, the disc is at focus. As the size of the Zernike disc is increased, the residual inequalities in the intensity at selected focus become more serious, and it may be inferred from figure 2 *b* that, in the case where the dimensions of the light-source can be neglected, the radius of a quarter-wave Zernike disc must

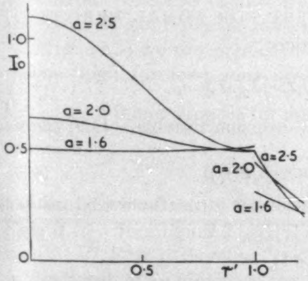


Figure 2 *a*. True mirror; intensities with phase-retarding disc at focus.

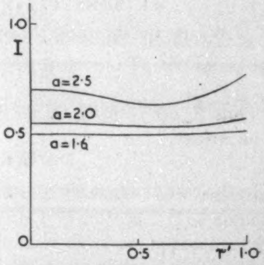


Figure 2 *b*. True mirror; with phase-retarding disc at preferred focal setting.

---

* Zernike (1934 a), p. 703.

† Calculated from (3.10); see next page.

not exceed about two-thirds of that of the Airy disc if a true mirror under the test is to appear uniformly illuminated.†

Zonal errors

A more convenient expression than (3.6) for $D(x', y')$ can be obtained when the surface under test suffers only from zonal errors, that is, when $E(x, y) = E(r)$ is a function of r alone. In this case $D(x', y') = D(r')$ is likewise radially symmetrical and is given by the equation

$$D(r') = E(r') + (e^{-i\alpha} - 1) \int_0^1 E(r) r \, dr \int_0^a J_0(\rho r) J_0(\rho r') \rho \, d\rho \qquad \ldots\ldots(3.9)$$

$$= E(r') + a^2(e^{-i\alpha} - 1) \int_0^1 E(r) K_0(ar, ar') r \, dr$$

$$= E(r') + a^2(e^{-i\alpha} - 1) \int_0^1 E(r) \frac{arJ_1(ar)J_0(ar') - ar'J_1(ar')J_0(ar)}{(ar)^2 - (ar')^2} r \, dr. \qquad \ldots\ldots(3.10)$$

Equation (3.10) reduces the problem of calculating the appearance under test of a mirror suffering from zonal errors of arbitrary form to the carrying out of a pair of numerical integrations for each value of r'.

§4. SENSITIVITY OF THE TEST FOR ZONAL ERRORS

We use for shortness the operational notation

$$Z_a E(r') = \int_0^1 E(r) r \, dr \int_0^a J_0(\rho r) J_0(\rho r') \rho \, d\rho. \qquad \ldots\ldots(4.1)$$

Thus $Z_\infty E(r') = E(r')$, by the Fourier-Hankel inversion formula. (3.9) can now be written

$$D(r') = E(r') - (1 - e^{-i\alpha}) Z_a E(r'). \qquad \ldots\ldots(4.2)$$

In the case of a true mirror tested with Zernike disc at focus, the wave-function is

$$E_0(r) = 1 \ (r \leqq 1)$$
$$= 0 \ (r > 1),$$

and $Z_a E_0(r') = I(r', a) = I_a$ can be calculated from (3.8). When the zonal errors are small, for example when they do not exceed 1/20 fringe, it is permissible to write

$$E(r) = 1 + iE^*(r) \quad (r \leqq 1)$$
$$= 0 \qquad (r > 1),$$

where $E^*(r)$ is real, and to neglect $(E^*)^2$. (4.2) then gives, if $r' < 1$,

$$D(r') = 1 + iE^*(r') - (1 - \cos\alpha + i \sin\alpha)(I_a + iZ_a E^*(r'))$$
$$= 1 - (1 - \cos\alpha) I_a + \sin\alpha Z_a E^*(r')$$
$$+ i[E^*(r') - I_a \sin\alpha - (1 - \cos\alpha) Z_a E^*(r')]$$

and the intensity

$$I = I(r') = |D(r')|^2 = 1 - 4\sin^2\frac{\alpha}{2} I_a(1 - I_a) - 2\sin\alpha [E^*(r') I_a - Z_a E^*(r')]$$

---

† Zernike (1934 b, p. 383) recommends the use of a disc between one-half and two-thirds of the size of the Airy disc for detecting astigmatism, and asserts that for detecting spherical aberration a disc more than twice as large as the Airy disc can be used. But his method of arriving at these estimates is not free from objection, since in the paper referred to no account is taken of the variation of intensity over a true mirror at preferred focus, on which chiefly depends, according to the present analysis, the upper limit to the useful size of the disc when small errors are under examination.

on discarding terms in $(E^*)^2$. That is to say

$$I - I_0 = -2\sin\alpha[E^*(r')I_a - Z_aE^*(r')], \qquad \ldots\ldots(4.3)$$

where

$$I_0 = 1 - 4\sin^2\frac{\alpha}{2}I_a(1 - I_a) \qquad \ldots\ldots\ldots(4.4)$$

is the intensity for a true mirror and

$$I_a = I(r', a) = \int_0^a J_0(\rho r')J_1(\rho)d\rho$$

can be calculated from (3.8).

In the case $\alpha = \frac{\pi}{2}$ of a quarter-wave retarding disc, (4.3) agrees with the corrected form of a result of Zernike.† When $r' > 1$, i.e. in the halo, (4.2) gives

$$\begin{aligned} D(r') &= -(1 - e^{-i\alpha})Z_aE(r') \\ &= -(1 - \cos\alpha + i\sin\alpha)(I_a + iZ_aE^*(r')), \\ I(r') &= |D(r')|^2 = 4\sin^2\frac{\alpha}{2}I_a^2. \end{aligned} \qquad \ldots\ldots(4.5)$$

Thus the appearance of the halo is not visibly changed by small zonal errors on the mirror. It is easy to show from (3.2) that the same conclusion follows when the errors are not restricted to be radially symmetrical.

From (4.3) we can predict in a general way the appearance under test of a shallow high (or low) zone on an otherwise true mirror. To fix ideas, let $\alpha = \frac{\pi}{2}$, $a = 2{\cdot}0$ and suppose for the present that the disc is at focus. Then (4.3) becomes

$$I(r') = I_0(r') - 2I_aE^*(r') + 2Z_aE^*(r').$$

On the right-hand side of this equation, $I_0(r')$ runs from 0·65 to 0·5 as r' increases from 0 to 1 (see figure 2 (*a*)), while $2I_a$ runs from 1·55 to 0·95. The second term, $-2I_aE^*(r')$, therefore contributes an intensity-decrease approximately proportional, at each point of the zone, to the height-excess of the surface at that point, and approaching 50% of the value of I_0 at a point where $E^*(r') = 1/4$, i.e. where the surface is $\frac{1}{25}$ fringe too high. The third term, $2Z_aE^*(r')$, describes the manner in which the light which has disappeared from the high zone is redistributed, fairly uniformly, over the whole mirror; as (4.5) shows, practically none of this light is sent into the halo. This redistributed light reduces the contrast with which the zone is shown up under the test, and its effect becomes more noticeable as the area of the zone increases.

A complication, the effect of which on the sensitivity of the test can be foreseen to be small without exact calculation, is that in practice the Zernike disc will not be set at focus but, in the present case, about $\frac{1}{50}$ fringe inside focus. It therefore seems safe to conclude that a fairly narrow zonal error of $\frac{1}{20}$ fringe should be easily visible on an otherwise true mirror and hence that the sensitivity of the test for errors of this type is of the same order as that of the Foucault test. As a check on this estimate, the intensity-distributions in a number of typical cases were computed from (3.10). A selection from the results is given in figure 3, which shows the appearance of high and low zones of $\frac{1}{20}$ fringe under the Zernike test at the preferred focus, for three different values $a=2{\cdot}5$, $a=2{\cdot}0$ and $a=1{\cdot}6$ of the disc-radius, and under the Foucault test with the knife-edge centrally set.*

---

† Zernike, 1934, p. 704, line 11, where the effect of the term in $Z_aE^*(r')$ is not taken into consideration; there is also a mistake in sign.

From figure 3 it appears that, for the detection of an error of this type, there is no appreciable difference in sensitivity between quarter-wave discs of radii $a=2{\cdot}0$ and $a=2{\cdot}5$, but that, with the larger disc, less light is diverted into the halo, and so lost for testing purposes. The sensitiveness is perceptibly decreased, however, if the value of a is reduced to 1·6. We conclude that, in testing local zonal errors, the sensitiveness of the test is nearly independent of the disc-size over the range $2{\cdot}0 \leq a \leq 2{\cdot}5$, and the upper limit to the useful size of the Zernike disc is accordingly determined by the consideration that a true mirror must not show too great inequalities in brightness when tested at the preferred focal setting.

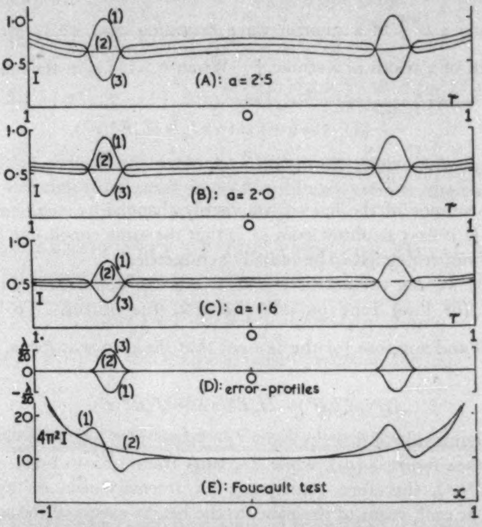


Figure 3. (A)-(C) Intensities under Zernike test with quarter-wave retarding disc, of radii $a=2{\cdot}5$, 2·0 and 1·6, at preferred focus: (1) low zone, (2) true mirror, (3) high zone. (D) Error profiles. (E) Intensities under Foucault test, knife-edge central; (3) is obtained on reversing (1) from left to right.

---

* The Foucault intensities were calculated from the formula

$$4\pi^2 I(x', 0) = \left| \pi E(x', 0) + i \int_{-1}^{1} \frac{E(t, 0)}{t-x'}\, dt \right|^2 \qquad (-1 < x' < 1).$$

REFERENCES

Burch, C. R., 1934. *Mon. Not. R. Astr. Soc.*, **94, 384**.
Linfoot, E. H., 1946. *Proc. Roy. Soc.*, **A, 186, 72.**
Zernike, F., 1934 a. *Physica*, **1, 689**; 1934 b. *Mon. Not. R. Astr. Soc.*, **94, 377**.

THE EFFECT OF THE ANGLE OF INCIDENCE OF THE EXPOSING LIGHT RAYS UPON THE RESOLVING POWERS OF PHOTOGRAPHIC MATERIALS

(Communication 1085H from the Kodak Research Laboratories, Harrow)

By J. M. GREGORY

Communicated by E. W. H. Selwyn ; MS. received 17 *May* 1946

ABSTRACT. It has been supposed that the angle of incidence of the exposing light would influence the apparent resolving power of a photographic material. Experiments have been carried out with a resolving power test object in contact with the photographic material, using illumination at different angles of incidence. The effect is almost indiscernible until angles of 40° are reached, at which it is of the order of 5 to 10 % reduction in resolving power with a medium-speed film. With fast film for aerial photography, the effect is barely observable. It is concluded that the drop in resolving power, observed in air camera negatives at points off axis, is not to any appreciable extent due to the above effect.

§ 1. INTRODUCTION

EXPERIMENTS have shown that the photographic resolution in camera negatives falls off at the edges of the field when the lens is focused on the film for the central part of the field. The reasons for this effect became of interest under a programme of research carried out for the Air Photography Research Committee of the Ministry of Aircraft Production, designed to provide the information necessary to advise on the improvement of the lens-film combination as used in air photography.

Selwyn and Tearle (1946) have discussed the effect in detail. It was suggested during discussions that the loss in resolution was due to the fact that, off the optic axis of the lens, the light enters the photographic emulsion obliquely whilst the processed film is, necessarily, viewed from the perpendicular. The effect expected is illustrated in figure 1 (*b*) in a highly idealized form. A screen, A A, having two apertures as shown, is illuminated by parallel light such that the photographic material is exposed at an angle to the incident beam. The shaded areas in the sensitive layer below it show the effect of the light on the layer as revealed by the subsequent processing. Examination from the perpendicular direction will no longer separate the two exposed areas, as would be the case when the exposing rays are normal to the sensitive layer (figure 1 (*a*)). This possibility was considered by lens-designers at one time to be an important factor setting a limit to off-axis resolution quite apart from considerations of the off-axis lens aberrations.

Dr. Stevens of these Laboratories has been able to demonstrate that effects of this nature can occur with a very transparent emulsion such as that of "Kodak Maximum Resolution Plates", but the work described here shows that there is no appreciable effect of this sort with negative materials of the type used in aerial photography.

It does not seem possible to devise a method for the direct determination of the effect of angle of incidence of the light rays upon photographic resolution which can give completely unequivocal results. Two methods, broadly speaking, are available. They may be termed the "Lens" method and the "Contact Printing" method respectively. The former method makes use of a very well corrected lens to throw an image of the test object upon the photographic material. It suffers from the disadvantage in the present application that as the photographic material is tilted to obtain the various angles of incidence required, the various parts of the test-object image not only go out of focus, but the image of the test object is also distorted. The latter method, which is the one used in this work, makes use of a special test object which is placed in good contact with the material under test, and exposed in much the same way as a contact print made from a photographic negative. The object is special in the sense that it has to be made much smaller than the test objects used for the lens method. In consequence, the photographic material used for making the test object must have the very highest possible resolution to record with sufficient perfection the minute images concerned.

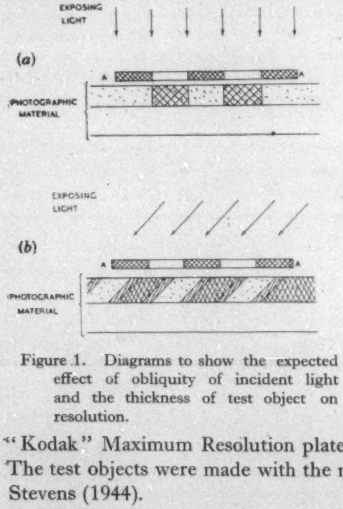


Figure 1. Diagrams to show the expected effect of obliquity of incident light and the thickness of test object on resolution.

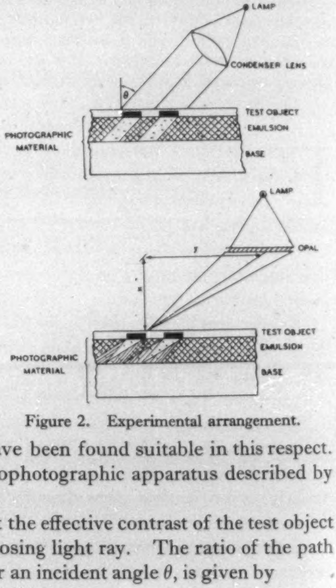


Figure 2. Experimental arrangement.

"Kodak" Maximum Resolution plates have been found suitable in this respect. The test objects were made with the microphotographic apparatus described by Stevens (1944).

The disadvantage of this method is that the effective contrast of the test object changes with angle of incidence of the exposing light ray. The ratio of the path length, L, to the test-object thickness, t, for an incident angle θ, is given by

$$L/t = \left[1 - \frac{1}{\mu^2}\sin^2\theta\right]^{-\frac{1}{2}},$$

μ being the refractive index (air to test-object layer).

Since the ratio is independent of the thickness, the effect cannot be eliminated experimentally by reducing the layer thickness. It can, however, be allowed for, and this will be dealt with later. Another difficulty, connected with the finite thickness of the opaque layer, is that the effective width of the various parts of the test object changes with the angle of incidence. This can be seen in figure 1 (*b*).

The difficulty can be overcome by making the opaque layer thin enough, since the effect depends upon the ratio of the layer thickness to the separation of the opaque parts of the individual charts of the test objects. For a given separation, therefore, the ratio can be made as small as we like by making the layer thickness sufficiently small.

For resolutions of 30 lines per millimetre (the order of resolution of air photographic materials) the separation between the bars of the test-object image is $17\,\mu$. Measurements, on the particular test object used in this work, using a microscope with an oil-immersion objective, show that the images are 3 to $4\,\mu$ thick; at 45° angle of incidence this would mean geometrically a reduction of the apparent separation by 3 to $4\,\mu$. Thus we might expect a purely geometrical reduction of the resolution by some 20% at 45°. This argument, however, has omitted consideration of the contrast profile of the test object as the angle of incidence increases. It can be seen that the path length of a light ray inclined to the normal going through the test object decreases as it approaches the edge. The effect of decreasing apparent separation, as the angle of incidence increases, must be largely compensated for by the parallel reduction in test-object contrast at the edges.

We have so far considered that the inclination of the ray takes place in a plane which is at right angles to the direction of the "bars" of the test chart. The light-scattering effects in the emulsion then take place across the area of material corresponding to the space between the bars of the test chart. We shall, however, be generally interested in the effect of inclination of the light ray regardless of the direction of the ray relative to the test-chart bars, for, in practice, details in the subject have a shape requiring good resolution in all directions.

The illumination of a point on a film given by a lens comes from all parts of the lens and forms a conical bundle of light rays pointing at the image point. One might expect, if resolution is at all effected by the angle of incidence of a light ray, that such a conical bundle of light rays could have a considerably greater effect than a beam of parallel light incident at the same angle as the axis of the cone. This point was therefore investigated experimentally by using an illuminated opal disc as a source of light for the contact exposures.

§2. EXPERIMENTAL

The diagrams of figure 2 show the two kinds of experiments which were made. In the first series of experiments, the pressure frame, holding the test object and test material clamped in good contact, was placed in the beam of light from a 5-inch diameter condenser lens arranged at such a distance from a 70-watt exciter lamp (which has a very compact filament) that the beam consisted of practically parallel light. By tilting the holder relative to the beam, a set of exposures was made at four different angles of incidence, namely 0°, 20°, 40° and 60° to the normal. In the second series of experiments, an evenly illuminated opal disc, of 6-inch diameter, was used as the source of illumination. It was placed parallel to the test object and distant 38 inches from it. The same four angles of incidence were obtained by shifting the frame parallel to itself by appropriate amounts, y, maintaining the distance, x, at 38 inches. Thus when disc and test object are opposite each other, the aperture of an equivalent lens would be f/6·3. This was chosen because it was one of the most common apertures in operational use for air-reconnaissance photography. In what follows, the first of these arrangements is referred to as yielding "parallel" illumination and the second as "diffuse".

The test object employed consisted of a row of units, each containing a range of sizes of Cobb charts. The unit is shown in figure 3. The Cobb charts are arranged in groups of a dozen all the same size but alternately at right angles to each other. The groups with the largest Cobb charts are placed on the outer edges of the unit. The size increments are 10% from group to group. By placing a neutral-density step wedge in front of the row of units a range of exposures could be obtained at the same time.

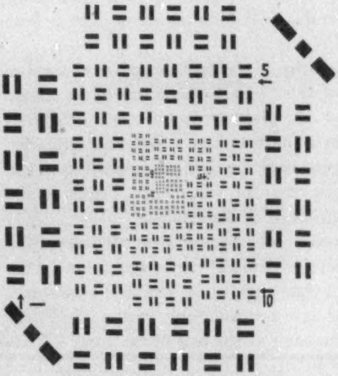


Figure 3. Test-object unit.

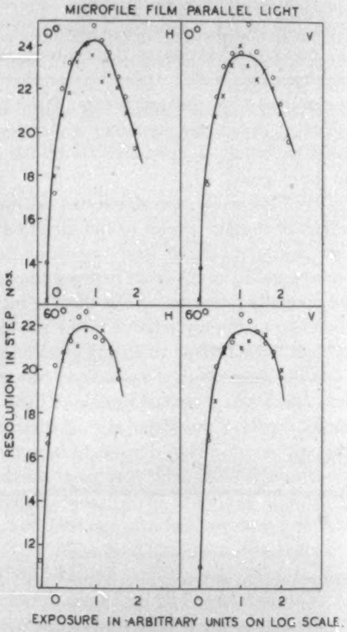


Figure 4. Typical resolution-exposure curves.

The design of the groups in alternate charts at right angles means that the resolution in two directions at right angles can be determined at the same time, by recording the results for the two directions separately. This means that some estimate of the general effect, regardless of the ray direction relative to the chart bar, can be obtained by averaging these two results.

The contrast of the test object is moderately low, there being a density of 0·38 between the light and dark lines of a chart, i. e. the light lines have a brightness of approximately $2\frac{1}{2}$ times that of the dark lines.

Two films of widely differing light-scattering properties were tested in this way. They were "Kodak Aero Pan Super-XX", a very fast photographic material used extensively for air photography, and "Micro-File" Panchromatic, a slow fine-grained material. Two strips of each film were exposed at each angle of incidence. The average thickness of the "Aero Super-XX" emulsion layer was about 25μ and of the "Micro-File" Panchromatic about 15μ.

All strips of each film were developed together in a machine giving high agitation of the developer. The "Super-XX" films were developed in Kodak developer D-19b for 8 minutes, and the "Micro-File" in Kodak developer D-16 for 4 minutes.

The processed strips were examined under a microscope at constant field brightness. Owing to the granular structure of photographic materials there is no sharp demarcation at which one group is resolved and the next smaller group is not resolved. Instead there is a fairly steady decrease in the number of individual charts resolved per group. The group at which 50% of the individual charts were resolved was therefore noted. A graph of exposure step *versus* group number was then drawn for each material and angle of incidence, the results for the repetition of each angle being plotted together.

§ 3. RESULTS

An example showing the resolution curve for "Micro-File" film at an angle of incidence of 60° with parallel light compared with the 0° curves is given in figure 4. The results for the two sets of test charts at right angles to each other are given separately under the arbitrary headings *vertical* (V) and *horizontal* resolutions (H), the former corresponding to the case when the ray inclination is in a plane at right angles to the chart bars. The *resolving power* of a material is defined as the maximum value of resolution on the resolution-log exposure curve. The average density at which the resolution-log exposure curve was a maximum was $D_{\max} = 1{\cdot}0$ for the "Aero Super-XX" film and $D_{\max} = 1{\cdot}3$ for the "Micro-File" panchromatic.

The resolving powers in the horizontal and vertical directions were all so nearly alike that they were averaged and the averages were then plotted as a function of the angle of incidence (figure 5 (*a*)).

§ 4. DISCUSSION OF RESULTS

The most noticeable features of the results were the lack of difference between the horizontal and vertical resolving powers and, to a less extent, the lack of an angle effect. It is only on reaching the extreme angle of 60° that a drop of resolution is encountered with the least scattering type of film, viz. "Micro-File". When the illumination is "diffuse" only the "Micro-File" results are influenced.

Before we can draw conclusions from these data, the effect of the test-object density with angle of incidence must be considered. Using the formula developed earlier for the increase in path length with angle of incidence, the percentage change in density may be computed. This is plotted in figure 5 (*b*) as a dotted line. Information on how such changes affect the resolution is meagre or inappropriate. The most appropriate data to hand are those of W. Romer (private communication) who used a well corrected quartz-fluorite apochromat by Hilger to throw the image of a series of test objects of various contrasts on to the same two photographic materials used in the present work. He found that in the region of test-object densities of 0·4, a 10% increase in density gave a 5% increase in resolution for "Micro-File" film and an 8% increase in resolution for "Super-XX". We may suppose with fair confidence that under the conditions of Romer's experiments, the effects found were mainly due to the material and that the lens did not interfere unduly with the results. By making proportionate allowances in the resolution figures for the change of contrast with angle of incidence (which applies equally to horizontal and vertical resolution components) we obtain figure 5 (*b*). We see that from the practical point of view the corrections are very small indeed.

Turning back to the discussion of the results then, we see that there is a small effect of angle of incidence on resolution, which is of little importance until very oblique angles are encountered. The effects are greater for "Micro-File" film both as regards the angle of incidence and in regard to the use of the diffuse illumination. This is readily explained in terms of the much smaller scattering of light in this material. The greater the light scattering in an emulsion the less the effect of angle may be expected to be. In the "Super-XX" layer the scattering is so great that the original angle of incidence of the light, or the question whether the incidence light is diffuse or parallel, hardly matters. An interesting conclusion may be drawn from the fact that the "Micro-File" film resolutions at 60° are the same for both diffuse and parallel light. The "diffuse" illuminator is such that the semi-angle of the illuminating cone gets progressively smaller as the angle of incidence increases, for the solid angle subtended by the opal disc at the test object decreases with increasing angle of incidence in the present set-up. This is due partly to the change in the projected size of the disc and partly to the increased distance from the test objects. Evidently, at 60°, the angle of the cone is so small that, as far as the "Micro-File" film is concerned, it is equivalent to "parallel" light.

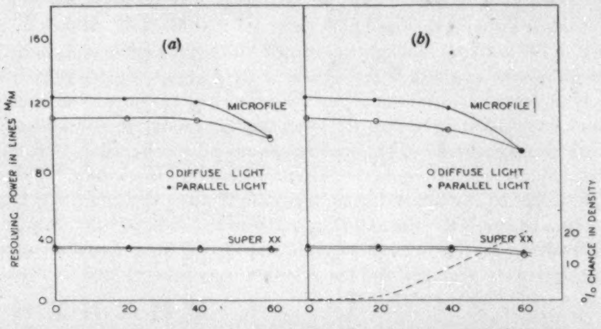


Figure 5. Graphs showing the effect of obliquity upon resolving power both before and after allowance is made for change in test-object density with obliquity. The percentage change in test object density with obliquity is shown by the dotted line.

We may conclude that there is an effect of angle to incidence of the exposing light upon the resolution of materials of the type used in air photography but that it is far too small to be of any importance in comparison with the drop in resolving power found when used in an air camera.

ACKNOWLEDGMENTS

I wish to express my thanks to colleagues engaged in the work of which this forms part for their interest during discussions, and to the Ministry of Aircraft Production for permission to publish this paper.

REFERENCES

Selwyn E. W. H. and Tearle, J. L., 1946. *Proc. Phys. Soc.*, **58**, 493.
Stevens, G. W. W., 1944. *Photogr. J.*, **81**, 108.

DISCUSSION

on papers by R. A. Scott entitled

(i) " The absorption of sound in a homogeneous porous medium."
(ii) " An apparatus for accurate measurement of the acoustic impedance of sound-absorbing materials."
(iii) " The propagation of sound between walls of porous material."

(This volume, pp. 165, 253 and 358.)

Dr. A. J. King. It may be helpful if I indicate briefly the circumstances which necessitated undertaking the three investigations described by the author in his group of papers. At the beginning of the war, we had in progress, in the Research Department of the company with which the author and I are connected, an investigation into the attenuation of sound in ducts such as are used in air-conditioning systems. This was to enable us to design ducts which would serve hotel bedrooms and ship's cabins without causing nuisance from noises due to fans and air turbulence. Measurements of sound attenuation were made in an experimental duct with various linings of porous material and an attempt made to correlate the results with the values predicted by the more or less empirical theories based on absorption coefficients put forward at that time. It was apparent that these theories were inadequate as they gave only a rough guide to the attenuation to be expected in any given case. When Morse published his theory it appeared to be much more comprehensive, but complete experimental verification was lacking.

At this point it became necessary to design a large engine-testing plant for war purposes, capable of testing jet engines at night without disturbing the neighbourhood. The magnitude and importance of the undertaking made it essential that the data used for design purposes should be reliable. Measurements were made of the intensity-frequency spectrum of the noise of an engine and of the ambient noise at night, thus enabling us to deduce the desirable attenuation-frequency characteristic of the silencing system. The existing investigation was therefore broadened, involving an extension of our acoustics laboratory to accommodate a more suitable duct and the development of apparatus, described in one of the papers, for the study of acoustic impedance. The agreement then observed between the experimental attenuations and those predicted by Morse's theory was much more satisfactory than in the case of the earlier theories, and so justified proceeding with the building of the test house. However, it was apparent that Morse's theory is incomplete, as it does not allow for the passage of sound along the material lining the duct. It was therefore necessary to study the mechanism of sound transmission in porous materials, and the author has shown that wave motion in them is similar to that in air. Unfortunately, the theory which the author has given, taking into account the passage of sound in the lining, does not lead to simple computation. However, he has shown that the difference between Morse's theory and the experimental results is in the direction indicated by his more complete theory, viz., that where Morse's theory predicts a high attenuation in a duct lined with loosely packed porous material, transmission along the lining is appreciable, resulting in experimental results lower than the predicted values. Where the predicted attenuation is lower, and the porous material denser, the difference is reduced. Thus, all three investigations were necessary for a complete appreciation of the problem.

Dr. G. H. Aston. I wish to ask to what extent the calculated attenuations, based on Dr. Scott's theory, for ducts with air spaces as used in practice have been compared with measured attenuations, and how far there is agreement.

Author's reply. In reply to Dr. Aston, the full range of ducts of interest in practice is very extensive and includes, at the one extreme, ducts which are split into channels of an inch or so in width by layers of absorbing material, and at the other extreme, ducts a few feet wide and lined with a few inches' depth of absorbing material. The first extreme lies close to the case referred to in figure 3 on p. 365. Thus, an examination of equation (15) on p. 362 shows that the attenuation coefficient calculated from my theory lies very close to that characteristic of propagation of sound of the same frequency in material of the lining in bulk. The experimental evidence put forward in figure **3** is in agreement with this deduction. At the other extreme of very wide ducts, the attenuation is usually very small (i.e. a few db./ft., and often much less). Comparison of equations (**15**) and (25) shows that the solutions of my equation agree under this condition substantially with those of Morse. In the course of a comprehensive investigation of the performance of the ducts, we have shown in our laboratories that for wide duct-spacings there is good agreement between measured attenuations and those calculated for the first propagational mode from Morse's (and therefore from my own) theory. We hope to publish a full account of this work in the near future.

There remains, however, the ducts for which the width of air-space and depth of lining are intermediate between the above extremes. Over this range we have considerable evidence of practical performance but practically no corresponding results calculated from my theory. The reason for the deficiency is that numerical computation of solutions of (15) in the general case are difficult. Prof. D. R. Hartree has suggested, in a very helpful discussion of the problem with the author, a manner in which the differential analyser might conveniently be used for the solution of the equations (7) and (9), and we hope to pursue this suggestion at some time in the future.

CORRIGENDA

Demonstration, "A Frank and Hertz critical potential experiment", by J. H. SANDERS (*Proc. Phys. Soc.*, **57**, 577 (1945)).

Page 578. In the wiring diagram of the oscillator (figure 2) the words "GRID" and "FILAMENT" should be interchanged.

"The transient flow of heat through a two-layer wall", by MIRIAM V. GRIFFITH and G. K. HORTON (*Proc. Phys. Soc.*, **58**, 481 (1946)).

Page 481, equation (1) : for d read ∂

Page 482, equations (4) and (5) : for d read ∂.

Page 482, last line :

$$\text{for} \quad \gamma=\frac{k_2S_2+k_1S_1}{k_2S_2-k_1S_1} \qquad \text{read} \quad \gamma=\frac{\sqrt{k_2S_2}+\sqrt{k_1S_1}}{\sqrt{k_2S_2}-\sqrt{k_1S_1}}.$$

Page 483, equation (14) : transpose factors, to read

$$(2H/p)\sqrt{D_1D_2/p}\,\{\lambda/(1+\gamma e^{2x\sqrt{p/D_1}})\}\exp\,[\alpha\sqrt{p}(D_1^{-\frac{1}{2}}+D_2^{-\frac{1}{2}})].$$

Page 483 equation (15) : insert a factor γ in the denominator of the term occupying the second line.

Page 483 equation (16) : to read

$$\frac{2H\sqrt{D_1D_2}\lambda}{\gamma p\sqrt{p}}\left(\frac{1}{1+\gamma^{-1}e^{-2x\sqrt{p/D_1}}}\right)e^{\sqrt{p}\left[x\left(\frac{1}{\sqrt{D_1}}-\frac{1}{\sqrt{D_2}}\right)-\frac{x}{\sqrt{D_2}}\right]}$$

Page 483 : at end of the second of the four lines occupied by equation (17), insert $)$.

OBITUARY NOTICES

MARIE PAUL AUGUSTE CHARLES FABRY

Prof. Charles Fabry was born at Marseilles on 11 June 1867. Though many of his working years were necessarily spent in Paris, he belonged to the South of France and, when opportunity allowed, returned there again and again. In a brilliant career at the Ecole Polytechnique he followed the example set by his grandfather, a pupil of Ampère. When later he became Membre de l'Institut, in the section of General Physics, he was following the lead given by two elder brothers, Eugène and Louis, who were already correspondents in Geometry and Astronomy, respectively.

Charles Fabry secured his *agrégation* in Physics in 1889 and his doctorate in Physical Sciences in 1892 with a thesis on the " Théorie de la visibilité et de l'orientation des franges d'interférence ". Two years later, after a spell of teaching physics in several Lycées, Fabry was called back to Marseilles as Chief Lecturer in Physics under Professor Macé de Lépinay. Here with Perot, already in charge of the teaching of industrial physics, he formed a very fruitful partnership in research and a number of important papers in interferometry and spectroscopy were published : a series on the application of fringes produced by silvered plates led to the well known Fabry-Perot interferometer described in the *Annales de Chimie et de Physique* for 1901. Perot left Marseilles for a post in Paris and was succeeded by Buisson ; then came a succession of important papers on absolute wave-length in the spectrum of iron and cadmium and on the establishment of a new system of standard wave-lengths in the laboratory and the Sun ; this met a need long felt not only in the laboratory but also in astrophysics. Although Perot, Benoît and Buisson collaborated in these papers, the sure hand of Fabry was present throughout.

In 1904 Fabry succeeded Macé de Lépinay in the Chair of Physics at Marseilles, and while continuing his teaching in optics made a great name for himself locally as a teacher of industrial electricity ; his lucidity, simplicity and humour drew overflowing crowds of students, engineers and workpeople to his Wednesday evening lectures. His own work in the electrical field lay in absolute measures, as in optics. But in the latter field along with absolute values Fabry also studied with the aid of the interferometer very small displacements of spectral lines, for instance between the spectra of the Sun at the centre and at the limb ; also he applied his objective interferometer to differences in the line-of-sight velocities in different parts of the Orion nebula and to a study of the temperatures of the luminous gases.

In 1914 the first World War broke into his life and he left Marseilles to work with a radiographic unit in the field; later he became, under Paul Painlevé, technical adviser to the Artillery in the National Office of Research and Inventions : this led to his being sent as head of a scientific mission to the United States, where he was already so well known and so popular that he was welcomed everywhere and given every facility in his task. His position as one of the world's leaders in optics was so clear that he was the inevitable choice as the first director of " L'Institute d'Optique " when it was founded in Paris in 1920. At the same time he became Professor of Physics at the Sorbonne with Prof. Cotton as his colleague. His work for the optical industry of France, both in the class-room and in the optical workshop and factory, is reflected to-day not least in the rise of a brilliant school of French astrophysicists, his devoted pupils, and in the development of new and effective optical instruments.

The Thomas Young Oration given to our Society in 1935 on " Vision in Optical Instruments " is typical of this side of Fabry's scientific life : it shows his interest in practical considerations as modifying the results of classical theory and his recognition of the part played by the eye when used as an element of the optical train.

Fabry's other address to our Society, the Guthrie Lecture, delivered in 1925 on " The absorption of Radiation in the Upper Atmosphere ", represents another of the main branches of his scientific activity—spectrophotometry and photometry. The lecture is mainly devoted to the evidence for the absorption of the solar radiation by ozone at high levels of the atmosphere : his study of this problem led to his presidency of the *Société Méteorologique de France*. Other spectrophotometric studies include an examination of the distribution of the solar spectrum in the range λ3150 A. – λ2900 A.

From boyhood, when he used to observe eclipses of Jupiter's satellites, Fabry was always interested in astronomy. His applications of interferometry to astrophysics have already been mentioned, but he also worked on direct astronomical photometry, tackling such problems as the brightness of the night sky and of the corona and the absolute magnitude of the Sun in the stellar magnitude scale. His report on " La photométrie astronomique et celle des physiciens " is a most illuminating exposition of the principles underlying the work of the astronomer and the physicist and of the precautions which they severally and jointly must adopt. Fabry also gave the George Darwin lecture to the Royal Astronomical Society in 1938, and his subject, " Interstellar Space ", links up with his general interest in radiations of all types.

Recognition of Fabry's qualities as a scientific leader came freely from his colleagues all the world over. He was Vice-President of the International Astronomical Union from 1932 to 1935, and again from 1938 to 1945, serving on several of its scientific commissions and being President for many years of its Commission on Instruments. He was President of the International Council of Scientific Unions from 1937 until he resigned for reasons of ill-health shortly before his death. He was a very welcome attendant at many international gatherings where his lucidity, good sense and quick humour were invaluable and widely appreciated. A great lover of his country, he was sadly aware in 1939 of the awful threat hanging over it and of his country's state of unpreparedness to meet the coming danger. The years that followed must have been sad ones for him, but happily he survived to see France liberated and taking the first steps towards recovery.

He received in 1918 the Rumford medal of the Royal Society, of which he became a Foreign member in 1931. In 1921 he received the Franklin Medal of the Franklin Institute and was elected as a member. He was made an Associate of the Royal Astronomical Society in 1915, received the Draper medal of the National Academy of Sciences in 1919, and was elected a member of the Royal Institution in 1920. He became an Honorary member of the Physical Society in 1926.

He died in Paris on 11 December 1945, after an illness of several months.

F. J. M. STRATTON.

EVAN JOHN WILLIAMS, F.R.S.

E. J. WILLIAMS was born in the heart of rural Wales in 1903, the son of a master stonemason. Educated at Llanwenog National School and then at Llandyssul County School, he took a scholarship which carried him to the University College, Swansea. Later he worked at Manchester under W. L. Bragg and then at Cambridge under Rutherford. After a year at Copenhagen with Bohr, he returned, first to Manchester and then to Chadwick at Liverpool. He was elected to the chair of Physics at Aberystwyth in 1936, and was elected into the Royal Society three years later.

He carried out both experimental and theoretical work on atomic and electronic collision processes. With the outbreak of war, his attention was turned to the immediate applied problems of the U-boat warfare, and, to quote Prof. Blackett's notice in *Nature* (**156,** 655), " It was largely due to Williams' keen analysis and powerful advocacy that an augmented and improved ' Bay offensive ' was staged in early 1943. The results, in terms of U-boats sunk, were not only in startling agreement with his predictions, but also were one of the important factors that led to the defeat of the U-boat campaign in 1943, and so made possible the serious planning of the invasion of Europe ".

(Based on material kindly made available by Prof. P. M. S. BLACKETT.)

IN the appreciation of E. J. Williams' life and work which appeared in *Nature* in December last, Professor Blackett has spoken of his brilliant contributions to the theory of collision problems ; he was also able, from first-hand knowledge, to do justice to the services which Williams rendered during the war in combating the U-boats. I did not know him so well in later years, but I can supplement Professor Blackett's account by my knowledge of his earlier career. He came to Manchester in the 'twenties as a young research student from University College, Swansea, with a warm recommendation from

Professor Evans. Our first impression was one of his blazing vitality and energy. I well remember the day of his arrival, because it happened to be the day of the annual rag between Owens College and the College of Technology, when a storming of each other's premises was a recognized procedure. Williams turned up in the laboratory next morning with his head swathed in bandages ; he had lost no time in entering into local politics. He was unconventional in the extreme. One usually knew where Williams was to be found in the laboratory from his habit of singing oratorios at the top of his voice when he felt like it, from sheer exuberance. To be driven by him in a car was a terrifying experience. He knew to a hair how close he could shave a tram, and would cut in between two advancing towards each other with a remark that it was really safer to do it that way because his brakes happened not to be working. He entered into everything he did with the same vast energy and enthusiasm. He and I collaborated in some work on the order-disorder phenomenon. It started at an afternoon colloquium, when I put forward some very tentative ideas in explanation of results on alloys which Bradley had been describing. Next morning Williams came to me with pages of calculations, embodying a well-developed mathematical treatment of the whole effect. It was so with all he did : he had an immediate and instinctive grasp of the essentials of a problem. Looking back now, I realize that we did not appreciate his full stature in those early days. He had a native genius for mathematical physics, brilliant but untutored, which needed for its full development contacts which we could not provide. He was at the same time no mean experimentalist, but his theoretical intuition made him impatient with the slow progress of experimental work ; his subsequent period at Cambridge was not as fruitful as it might have been because the division of energy between his experimental and theoretical interests hindered his deploying his full powers. Great though his contributions to science have been, one cannot but believe that had he lived he would have risen to still greater heights, and gathered a notable school of researchers round him in his own department. His death, as Professor Blackett says, leaves a sad gap in the ranks of our theoretical physicists. It is tragic that a man of such exuberant vitality and brilliant originality should have died so young.

W. L. BRAGG.

JOHN LOGIE BAIRD

J. L. BAIRD, a Fellow of the Physical Society since 1927, died at his home at Bexhill on 14 June 1946 after a few months' illness, previous to which he had been actively engaged in experimental research in television in the laboratories of his own company.

Baird was the son of a Scottish minister and received his scientific education at the Royal Technical College, Glasgow. He first started on a business career, but later abandoned this to devote himself to an experimental study of the problems of television, which occupied him for the remainder of his life. Working under conditions of great difficulty due to limited means and apparatus, Baird correctly appreciated the basic problem of television, which consists in the provision of means for scanning an image by subdividing it into tiny elements, and transforming the resulting light variations into electrical impulses for transmission by line or radio to the receiver, where the impulses are converted back into light for the reconstruction of the picture. Using a mechanical scanning device and very intense illumination of the picture to be transmitted, Baird gave, in January 1926, what is claimed to be the first practical demonstration of television. This was followed in May 1927 by a demonstration of reception in Glasgow of picture images transmitted from London, and in February 1928 by the successful transmission of television across the Atlantic.

After these tests Baird Television Ltd. was formed to develop the technique for practical application, and the first step towards the inauguration of a public television service in Great Britain was taken in 1929, when the B.B.C. decided to give the above company facilities for experimental transmissions through the medium-wave London broadcasting station. Later the system was improved and adapted to operate on ultra-short waves, and towards the end of 1936 a public service was started from Alexandra Palace and comparison tests were conducted with the Baird and Marconi-E.M.I. systems, the latter being the one ultimately selected and now in use at the recently re-opened London Television Station.

During all these years, and until a few months before he died, Baird continued to work steadily towards the improvement of the scope and possibilities of television and its presentation. He had early taken advantage of the possibilities of the cathode-ray tube for reception, and also demonstrated the optical projection of the television pictures on to a cinema screen The possibilities of introducing colour and stereoscopic effects also aroused his interest, and in 1944 he gave a demonstration of his recent achievements in the reception of television in colour by a method which avoided the need for revolving disks and lenses.

To some extent Baird's early work suffered some discredit and lack of recognition as a result of the exaggerated publicity and premature claims given to it by his supporters, but Baird himself was a shy and modest man, and there is no doubt that his pioneer achievements were a great incentive to other workers, and contributed materially towards the success attained by radio engineers and physicists, placing Britain in the forefront in this fascinating application of electromagnetic waves. R. L. SMITH-ROSE.

GERALD SYDNEY FAWCETT

G. S. FAWCETT won for himself a unique position in colour physics. For nineteen years, until his death after a short illness on 8 November 1945, he was managing director of The Tintometer Ltd., the firm founded by his grandfather, J. W. Lovibond. In this position it was his constant desire to see his grandfather's invention used and developed to the benefit of his fellow men. Recent developments which enabled the Tintometer to be used in medical science were a source of particular satisfaction to him.

Born in Salisbury in 1896, he was educated at the Modern School, Salisbury, and at Christ's Hospital. Between 1915 and 1918 he served in the R.A.M.C. and in the Labour Corps. He then trained and practised for a short time as a quantity surveyor. In 1925 he married Marion, daughter of Mr. and Mrs. Oscar Wheater. She and three children, Paul, Janet and Joe, survive him.

As a boy he took a great interest in his grandfather's work, and he gained his wide knowledge of colour physics through individual study and practical experience rather than from systematic instruction. He became a Fellow of the Optical Society in 1926 and of the Physical Society in 1932. He joined the Colour Group when it was formed in 1940, and was a member of its committee at the time of his death. He was joint author with Prof. R. H. Stoughton of *Chemical Testing of Plant Nutrient Solutions*.

To know Gerald Fawcett was to know a friend, and what a good friend he was ! It was a cruel fate that took him from us so soon, but our memories of him remain an inspiration. R. K. S.

DANIEL EVAN JONES

By the death of D. E. JONES the Physical Society has lost yet another of the fast-dwindling band of Fellows who joined its ranks in the first fifteen years of its existence. He was elected to the Society in 1886.

After a period of study at the University College of Wales, Aberystwyth, and at Owens College, Manchester, under Sir Henry Roscoe and Balfour Stewart, he took his B.Sc degree in 1885 at London University with first-class honours in chemistry. For some years he held the chair of physics at University College, Aberystwyth, but, feeling that his gifts lay rather in the direction of administration, he left the teaching profession for the posts of Director of Technical Education to the Staffordshire County Council and Regional Inspector under the Department of Science and Art. He was the author of an elementary textbook on Heat, Light and Sound, and of a set of examples in physics ; these books had a considerable vogue some forty years since.

But Jones's major contribution to physics, and one by which he will be most gratefully remembered, consists in his lucid and scholarly translation of the works of Heinrich Hertz. In 1893 he published, under the title of *Electric Waves*, translations of Hertz's epoch-making papers. The volume was introduced by a very characteristic preface, which is still of historical interest, from the pen of Kelvin. In 1896 appeared a translation, made in conjunction with G. A. Schott, of Hertz's *Miscellaneous Papers*, and in 1899, with the assistance of J. T. Walley, a translation of Hertz's *Principles of Mechanics*. The two last-named volumes contained most valuable biographical and critical introductions by Lenard and by Helmholtz, and the three volumes were of great value in making the work of Hertz accessible to British physicists. They are the translator's most enduring monument.

Jones appeared occasionally, after his retirement, at meetings of the Society, and is remembered as a kindly man, a close observer and a shrewd critic. A. F.

REVIEWS OF BOOKS

Dictionary of Science and Technology in English-French-German-Spanish, by MAXIM NEWMARK. Pp. viii+386. (London: Sir Isaac Pitman and Sons, Ltd., 1945.) 30*s*.

The author is in the Modern Language Department of the Brooklyn Technical High School, and has experience both of teaching technical students and of translation. He began his book as notes from experience, and then systematically extended it by compiling word lists from the indexes to textbooks and by examination of other technical dictionaries.

It is arranged with the English words in alphabetical order, followed by the three equivalents. Each such English word has a number, and there are indexes of French, German and Spanish words, giving reference to the initial letter and number of the corresponding English word. Each continental word has its gender set out in the main table, a feature much to be commended. It is a little difficult, however, to see precisely the class of user to whom the dictionary is most intended to appeal. A translator from English to one of the other languages finds the words he requires at once, with their genders, but he hardly needs the non-technical words which are so copiously supplied (across, to act, broom, brush, electric, index, nature, under, zig-zag and zinc). The translator, and indeed the general reader who is competent to undertake the perusal of matter in French, German or Spanish, could equally well do without most of these words, or at any rate could take them from an ordinary dictionary in the rare cases where a peculiarity of use of prepositions is in question, but he can hardly need the words which are alike in all languages. Thus we have on a single page

Macadam	macadam	Makadam	macadám
Machine	machine	Maschine	máquina
magnesia	magnésie	Magnesia	magnesia
magnesite	magnésite	Magnesit	magnesita
magnesium	magnésium	Magnesium	magnesio

and on another

cupreous	cuivreux	Kupferhaltig	cúprico
cyanamide	cyanamide	Zyanamid	cianamido
cyanide	cyanure	Zyanid	cianuro
cycle	cycle	Zyklus	ciclo
cycloid	cycloïde	Zykloide	cicloide,

though it is true that on this page we also find the entry " cyclotron ", " same in all languages ".

As to the actual coverage, there are 10,000 entries and the book has a vast number of workshop terms, a great deal of physics and chemistry, and practically no biology. The word " bee " does not appear, nor does " owl ", but " wheat " and " oats ", " birch " and " beech " do, though " barley ", " corn " and " maize " do not. Notable omissions are " bulldozer " and " fission ". There are a few surprises, most of which reflect rather the difference between English and American than any error on the part of the author. Thus " Adobe " translates into French " Argile " and German " Tonerde ", and " clay " into " argile " and " Lehm ", whereas " mud " is " Bourbe " and " Schlamm "; under " navvy " we read " same as *steam shovel* ". One meaning given for *airline* is " à vol d'oiseau ", which in this country could only be expressed as " bee line " or " as the crow flies ". " Police " appears, but not " Policeman ", so that " Agent " in the French index misses one important meaning, and the familiar " Schupo " does not appear at all. Finally, without any suggestion that there is an omission or error, we may sympathize with the French on learning from this dictionary that every time they wish to speak of a radio beam they have to write or say " rayon radiogoniométrique ".

J. H. A.

Relaxation Methods in Theoretical Physics, by R. V. SOUTHWELL. Pp. vi + 248. (Oxford: The Clarendon Press, 1946.) 20*s*.

The value of the relaxation methods introduced some years ago by Professor Southwell is by now well known, and it is generally realized that they are not confined to problems in statics. In the present volume, it is shown that these methods may be extended to deal with the solution of boundary-value problems associated with partial differential equations in two independent variables. For this purpose, the equation is replaced by its equivalent in finite differences, and a network covering the region of space concerned is set up, on which the values of the dependent variable can be computed by a method in which the boundary conditions are complied with, and the equation is satisfied by adjustments similar to those used in the more familiar relaxation method.

The basis and the methods used are both explained clearly, and many solved examples are given, so that the reader, even if disinclined to work out the whole problem for himself, may verify quite quickly that the solution given does in fact satisfy the required differential equation. As was remarked in the review of the earlier book, *Relaxation Methods in Engineering Science* (and more neatly by a reviewer quoted with approval by the author), it is essential for a reader who really wishes to grasp the subject to perform a few computations himself. Even without this, however, it is possible to appreciate the value of the method to some extent, and to admit that when it can solve problems of which the solution has hitherto been sought in vain, it must be worthy of attention. Several such problems are dealt with, including the flow of liquid under gravity to form a waterfall.

With such a record of success, and with an exposition which would in any case predispose a reader in its favour, it seems a pity that the author has thought it necessary to write about the method in terms which would suggest that there is a conspiracy against it, and that it is a rival (instead of a supplement) to orthodox methods, the practitioners of which are attempting to suppress it. Thus we are told that the aims of orthodox analysis of a physical problem are to obtain general and exact solutions, and it is implied that relaxational methods have different aims. Now, in so far as the new methods solve only one case of a given problem (for example, one geometrical parameter may have to be fixed), it is surely indisputable that a general solution would be preferable. But as regards accuracy, it seems to the present reviewer that the relaxationists and the followers of orthodox analysis both have the same aim—to solve the problem within pre-specified limits. The former do it numerically, the latter may give, for example, a convergent expansion, of which as many terms are worked out as may be needed, or they may give an asymptotic expansion, in which case care is taken to see that the possible error is less than the pre-specified figure for those values of the variables which are in question. Again, a good deal is said about the meaning attached by the theoretical and the practical man to the word " function "; but in fact nothing which is said here would surprise a reader of orthodox treatises, particularly of finite differences or calculus of observations, save in the section on basic theory of conformal transformation, where " function " suddenly takes on the meaning assigned by the rigourist to " analytical function ", and does not have the meaning so carefully set out for it by the author.

This " Heaviside complex " apart, the book is a model of good exposition, on a subject of real importance. It is beautifully produced, and the publishers deserve commendation ; their task was not easy, for in order to give the solutions of problems in a legible manner, it has been necessary to include about a dozen folding plates, as well as to set up a great deal of mathematical and tabular matter. J. H. A.

Electric Discharge Lighting, by F. G. SPREADBURY. Pp. viii + 136, 122 figures. (London: Sir Isaac Pitman and Sons, Ltd., 1946.) 15*s*.

Rapid progress has been made in the development of electric discharge lamps during the past 10 or 15 years, and a great variety of types, from simple negative glow lamps dissipating only a few watts to high pressure mercury vapour lamps dissipating several kilowatts, have been developed. In some of these lamps the discharge itself is the light source while in others the light is obtained from a layer of luminescent material excited by the ultra-violet radiation from the discharge which contributes only a negligible amount of direct light. The importance of these lamps lies mainly in the fact that the electric discharge permits of much greater possibilities as regards luminous efficiency and control of colour than does the tungsten filament lamp.

In the book under review the author has, somewhat ambitiously, attempted to deal with all types of discharge lamps in the space of some 130 pages. Following an introductory chapter describing certain theoretical aspects of radiation and electric discharges, there are three chapters, of some 60 pages in all, dealing with the lamps themselves and including a section on luminescence. The lamps described are those developed by the leading lamp makers in this country and no attempt has been made to deal with American electric discharge lamp practice. Although the book claims to deal with all types of discharge lamps in current use there is no mention of the important and interesting compact source mercury vapour lamps, which were largely developed in this country. Two chapters are devoted to circuit and gear matters, while there is a somewhat irrelevant concluding chapter on *Technical Applications of Discharge Lamps* ; it comes as something of a surprise to find the use of the gas discharge triode described in this chapter.

It is not clear for what type of reader the book is particularly intended : workers in the field will hardly find anything new or stimulating in what is, after all, little more than an assembly of well-known information. The lighting engineer is likely to be disappointed in the book for, despite its title, lighting problems are simply not discussed. It appears probable, however, that the book is not addressed primarily to either of these categories, but rather to the general reader who wants less specialized accounts of particular fields of activity, and to this class of reader it is recommended.

The book is excellently produced and printed, and the author has managed to assemble a considerable amount of material within the compass of a very compact volume. The diagrams and illustrations are excellent and the reproduction of oscillograms of voltage and current wave forms is particularly pleasing. A useful feature of the book is the collection of tables and curves giving performance and dimensional characteristics of British discharge lamps. It should be realized, however, that such a collection of data is liable to get out of date quickly—indeed some of the information given in the tables is already out of date—and it is always advisable to consult the lamp manufacturers for present-day information relating to their products.

In writing a book on practically any technical subject, the problems of acknowledging the contributions of other workers in the field naturally arises. The author of the present book solves the difficulty by avoiding acknowledgments. Without presuming to criticize this procedure, it may perhaps be suggested that the value of the work would have been greatly enhanced by the inclusion of a fairly full bibliography, which, in addition to assisting the reader who wants to probe deeper into the subject, would serve to indicate the source of much of the work described in the book. No doubt this is a matter which can be considered in connection with subsequent editions. H. G. JENKINS.

Mathematical Tables. Part-volume A, Legendre Polynomials, by the Committee for the Calculation of Mathematical Tables. Pp. 42. (Published for the British Association at the University Press, Cambridge, 1946.) 8*s*. 6*d*.

The volumes issued by the British Association Tables Committee ushered in the new era of plentiful, and reliable, tables, and these volumes are now well known to all who work in applied mathematics. In the past, the volumes have been of comparatively large size, and usually devoted to one type of table. The need has now been felt for the publication of various smaller tables, and to avoid waiting until enough disconnected material has been accumulated, the Association has adopted the device of issuing part-volumes.

The issue now under review contains the Legendre polynomials $P_n(x)$ regular at the origin. When the work was started, there were three tables in existence—the famous one of Perry up to $n=7$, and two up to $n=8$, all for $x=0$ to 1. The present table carries n up to 12, and in this range of x all are correct to at least 7 figures. There were at the inception of the project no tables beyond $x=1$, though in the interval several have appeared (Tallqvist, 1937 ; Prevost, 1933 ; Dwight, 1945). In the present volume, x is taken to 6 for all the values of n given, and to 11 for the first six values of n.

The tables are thus a valuable complement to those previously available. They are well printed and carry a clear and useful introduction by L. J. Comrie, whose name on a book of tables is almost a guarantee that there are no errors. J. H. A.

Mathematical Tables. Part-volume B, the Airy Integral, by J. C. P. MILLER. Pp. 56. (Published for the British Association at the University Press, Cambridge, 1946.) 10*s*.

The complete solution of the equation $y''=xy$ is $y=A\,\mathrm{Ai}(x)+B\,\mathrm{Bi}(x)$, or in an alternative form, $y=CF(x)\sin[\chi(x)+\epsilon]$. With a linear change of both variables, the function $\mathrm{Ai}(x)$ gives the integral $\int_0^\infty \cos\tfrac{1}{2}\pi(w^3-mw)\,dw$, whilst both Ai and Bi are closely related to modified Bessel Functions of order $\pm 1/3$. In the present part-volume, Ai and its first derivative are given for $x=-20$ to $+2$ at intervals of 0 01, and Bi from -10 to $+2{\cdot}5$ at intervals of 0·1. The logarithms and logarithmic derivatives of both are also given, the former from 0 to 25 at intervals 0·1 and thence to 75 at unit intervals, the latter from 0 to 10 at intervals of 0·1. The first 50 zeros and turning values of Ai and the first 20 of Bi are also given, together with tables of F and χ and the corresponding functions for the derivatives.

In addition, as we have come to expect from the B.A. Tables, there is a useful introductory treatise on the mathematics of these functions, with an elaborate account of the computational methods used. J. H. A.

Electrons in Action, by J. G. DAUNT. Pp. 151. (Sigma Books, 1946.) 6*s*.

In the preface of this book the author states that his purpose is to give the non-scientific reader an understandable account of electricity in its various aspects without recourse to mathematical symbols or formulae. He succeeds in the avoidance of technical algebra, except for the mention of Ohm's Law in the appendix, and here the use of the symbol C for current is perhaps a little unfortunate as being out of accord with modern practice. In making his subject intelligible to the ordinary reader Dr. Daunt has certainly spared no effort to find apt analogies with which to illustrate his points ; his style throughout is interesting, though as a whole it lacks the fascination of the Mr. Tompkins books.

The field of electro-technics covered by this book is very wide, the author commencing with a simple exposition of the electron theory of matter and, with these fundamental ideas established, the reader is introduced in turn to the phenomena of conduction of electricity, electro-magnetism and electric heating. Later chapters deal with thermionic-emission, X rays, photo-electricity, radar, cosmic rays, etc., and finally the author introduces the idea of the dualism of matter and Heisenberg's *principle of uncertainty*.

The work contains a number of photographic plates appropriately chosen, together with a large number of line drawings which, though often quite original in conception, are in some cases marred by poor execution. This book should prove a most useful introduction to the subject for non-scientific school-pupils and adults alike, and for those readers who are sufficiently stimulated to read further, Dr. Daunt has thoughtfully inserted, in the appendix, a list of books of graded difficulty. R. W. B. S.

Alternating Current Measurements at Audio and Radio Frequencies, by DAVID OWEN. Pp. 120+vii, with 80 diagrams, Second Edition, revised. (London: Methuen and Co., Ltd., 1926.)

This book is a second and revised edition of a popular monograph of the Methuen series. In it the general features of the original have been retained, the opening chapter containing the requisite A.C. theory for the clear understanding of the various bridges, potentiometers and resonance circuits described in succeeding chapters. The experimental procedure and the calculation of results from typical observations are given with each method described, the clarity of the explanation in the text being enhanced by means of line diagrams. The chapter dealing with the conditions of accuracy in bridge measurements at low frequencies is most instructive and useful, and in fact the only note of criticism is in regard to the section on radio-frequency measurements. The author has neglected here an opportunity to bring the text right up to date, for contrary to his remarks, bridge methods at these high frequencies are now actually in commercial use.

R. W. B. S.

Optique Instrumentale, by G. A. BOUTRY. Pp. x+539. (Paris: Masson et Cie, 1946.)

Professor Boutry's charming book represents the course of instruction on the theory of optical instruments given by him to students at the Institut d'Optique. It is not concerned with the problems of optical design, but first examines the fundamental theorems on which all sound optical designing is based, and passes on to consider the construction and use of representative instruments of modern types.

As the author explains in an interesting preface, his aim is to present a clear picture of the subject—still living and growing—rather than to adhere to any pedantic limitations, although it must be said that little exception can be taken to the general foundation. The mathematics takes, however, a generally subordinate position, the main effort being towards clarity of exposition in which the use of frequent clear diagrams will appeal to most students. The graphical presentation of the changes in spherical aberration of a refracting surface according to conjugate distance can be mentioned as a praiseworthy sample of the use of the diagrams to convey in one glance information which can only come from hard work with a formula.

There is a very useful and up-to-date chapter on the Eye and Vision in relation to the use of optical instruments. The well-known researches of Arnulf and others at the Institut d'Optique are freely drawn upon for material here.

If there can be any general criticism (apart from detail) one might feel this general picture of the edifice of optical knowledge is almost too well-drawn. There are only " one or two " references to original papers, and even references to proper names are relatively few, and mostly restricted to classical authors.

In the discussion of the theory of the microscope, for example, the reader would get no impression that this subject has been a battle-ground of strong controversy, and the treatment in the book still seems to lead to erroneous ideas regarding the proper function of the condenser in microscopy ; it introduces once more (though not in the same words of course) the old error of critical illumination.

The time is passing, perhaps, when a student expected a textbook to be verbally inspired on every detail. While directness of statements is generally to be welcomed, it should not prevent an author from indicating the points where the discussion is still incomplete and further enquiry is necessary. A book may be like a perfectly painted picture by van Eyck ; one leaves it in admiration—but an unfinished Michael Angelo may provide more stimulation to the imagination. In spite of the author's professed aim in the preface, he has not always been successful in indicating the " lines for development ", though a great many comments throughout the book will undoubtedly prove stimulating and helpful to any optically-minded reader ; probably he feels that the " lines for development " belong more to the province of the designer than to the general student of optical instruments, but even so, the attractive possibilities of systems with aspheric surfaces—including the Schmidt camera—might have been expected to receive a fuller discussion.

Professor Boutry has asked that a minor defect (on page 217) in the treatment of systems not possessing full axial symmetry should be pointed out : a statement appropriate to a single surface has been inadvertently generalized to cover a complete system. On the whole, however, the book seems very free from errors.

Its sign convention will not, perhaps, appeal to the majority of British readers; it is not in accordance with the systems which have been most favoured in this country. There must be very few, however, who could peruse the book without adding most usefully to their optical knowledge.

L. C. MARTIN.

INDEX

PAGE

Aberrations of optical systems, Algebraic analysis of the higher-order 545
Absorption and reflection of radiation of 9-cm. wave-length, Experimental investigation of the 265
Absorption band-systems of SiS, SiSe and SiTe 538
Absorption in optical glass, Note on measurements of 472
Absorption of sound in a homogeneous porous medium; and discussion 165, 775
Acceleration (vertical) and variation of pressure with depth in a liquid 118
Acoustic impedance of sound-absorbing materials, Apparatus for accurate measurement of the; and discussion 253, 775
Aerial, Fluctuation noise in a receiving 313
Aircraft camera lenses, The performance of 493
Alloy of iron and silicon, Hysteresis and eddy losses in single crystals of 21
Alloys, Magneto-resistance of high-coercivity 153
Angle of contact of mercury, Hysteresis of the 120
Aspheric plates, Decentred 65
Astigmatism formulae, A transformation of known 663

Baird, J. L. (Obituary notice) 779
Band, W.: Mendelssohn's z-particles and the Bose-Einstein statistics 302
Band systems of N_2 292
Band-systems of SiS, SiSe and SiTe 538
Band-systems of SnSe and SnTe 707
Bands of SiS, rotational analysis 606
Barr, G.: Errors in viscometry due to surface tension 575
Barrer, R. M.: Diffusion and thermal conductivity "constants" in non-homogeneous media, and in media where these "constants" depend respectively on concentration or temperature 321
Barrow, R. F.: Rotational analysis of some of the near ultra-violet systems (D–X) of SiS 606
Barrow, R. F., *see* Vago, E. E.
Bates, L. F.: The magneto-resistance of high-coercivity alloys 153
Baxter, S.: The thermal conductivity of textiles; and discussion 105, 305
Birge-Sponer extrapolation, The determination of dissociation energies by the 525
Bloch, A.: On methods for the construction of networks dual to non-planar networks 677
Bose-Einstein statistics and Mendelssohn's z-particles 302
Brightness-colour space, A modified Helmholtz line-element in 41
Buchdahl, H. A.: The algebraic analysis of the higher-order aberrations of optical systems. Tangential aberrations of a system of coaxial spherical refracting surfaces 545
Burgess, R. E.: Fluctuation noise in a receiving aerial 313

Calculation of optical constants 704
Caldin, E. F.: Determination of the initial temperature of a cooling total radiator from measurements of the spectral distribution of the energy emitted during cooling; and discussion 350, 356
Caldin, E. F.: Light emission during cooling of a Planckian radiator; and discussion 341, 356
Caldin, E. F.: Ratio of the new international lumen to the lightwatt; and discussion 207, 356
Calthrop, J. E. (Obituary notice) 332
Camera lenses, aircraft, The performance of 493

PAGE
Campbell, N. R. and Hartshorn, L.: The experimental basis of electromagnetism: The direct-current circuit 634
Carbon-monoxide, Photochemical decomposition of; and discussion . . . 701
Cathodo-luminescence: I–III 369, 383, 392
Cavity magnetron 247
Cell dimensions, Accurate determination of, from single-crystal x-ray photographs . 200
Centimetric wave-lengths, Radar echoes from the sea surface at 717
Chambers, Ll. G., *see* Temperley, H. N. V.
C.I.E. trichromatic system, Colours of total radiators expressed on the . . 1
Cinematograph projector, Optical problems of the rotating-prism 598
Cockcroft, J. D.: Second Rutherford Memorial Lecture 625
Colours of total radiators expressed on the C.I.E. trichromatic system . . . 1
Combustion, Thermodynamic properties of the products of high-pressure . . 737
Conductivity (thermal) of textiles; and discussion 105, 305
Cooling of a Planckian (total) radiator; and discussion 341, 350, 356
Corner, J.: Thermodynamic properties of the products of high-pressure combustion 737
Corrigenda 340, 776
Cosslett, V. E.: Variation of resolution with voltage in the magnetic electron microscope 443
Cripps, Sir Stafford: Address delivered at the opening of the Thirtieth Exhibition of Scientific Instruments and Apparatus 125
Cruickshank, F. D.: A system of transfer coefficients for use in the design of lens systems: VI 296
Crystal structure of double oxides of the perovskite type 133
Crystal x-ray photographs, Accurate determination of cell dimensions from single, . 200

Davies, H. and Macfarlane, G. G.: Radar echoes from the sea surface at centimetre wave-lengths 717
Debye-Scherrer lines produced by divergent x-rays, Integral breadths of . . 401
Decay characteristics of luminescent solids, The rapid determination of . . 289
Decentred aspheric plates 65
Diatomic molecules, Electronic states and potential curves of 695
Diffusion and thermal conductivity "constants" in non-homogeneous media, Measurement of 321
Diffusion of light by magnesium oxide 408
Dissociation energies, The determination of, by the Birge-Sponer extrapolation . 525
van Doninck, W., *see* van Itterbeek, A.

Eddy losses and hysteresis in single crystals of an alloy of iron and silicon . . 21
Electro-magnetism, The experimental basis of 634
Electron microscope, The variation of resolution with voltage in the magnetic . 443
Electron optical systems, Complete computation of 30
Electronic states and potential curves of diatomic molecules 695
Emission of light, during cooling of a Planckian radiator; and discussion . 341, 356
Energy distribution and initial temperature of a cooling total radiator; and discussion 350, 356
Errata 340, 776
Exhibition of Scientific Instruments and Apparatus, Address delivered at the opening of the Thirtieth 125

Fabry, C. (Obituary notice) 777
Farquhar, Margaret C. M. and Lipson, H.: The accurate determination of cell dimensions from single-crystal x-ray photographs 200
Fawcett, G. S. (Obituary notice) 780
Fleming, J. A.: Geomagnetic secular variations and surveys 213
Fluctuation noise in a receiving aerial 313
"Fly's-eye", Construction and use of, for assisting x-ray structure analysis . . 306
Ford, L. H. and Oliver, R.: An experimental investigation of the reflection and absorption of radiation of 9-cm. wave-length 265

PAGE

Fundamental frequency of vibration of rectangular wood and plywood plates ; and discussion 78, 487

Gaydon, A. G. : The determination of dissociation energies by the Birge-Sponer extrapolation 525
Gaydon, A. G. and Herman, Renée : Band systems in the spectrum of nitrogen 292
Gels, rigid hygroscopic, Some thermodynamic relations of 585
Geomagnetic secular variations and surveys 213
Gerö, L., *see* Schmid, R. F.
Grant, Kerr : The dependence of pressure variation, with depth in a liquid, on the vertical acceleration of the liquid : demonstration 118
Gregory, J. M. : The effect of the angle of incidence of the exposing light rays upon the resolving powers of photographic materials 769
Griffith, Miriam V. and Horton, G. K. : The transient flow of heat through a two-layer wall 481
Griffith's theory of rupture, Extension of, to three dimensions 729
Guggenheimer, K. M. : New regularities in vibrational spectra 456

Harding, H. G. W. : Colours of total radiators expressed on the C.I.E. trichromatic system for the temperature range 0–1–660 mireds (C_2)=14384·8) 1
Harrison, V. G. W. : The light-diffusing properties of magnesium oxide 408
Hartshorn, L., *see* Campbell, N. R.
Hearmon, R. F. S. : The fundamental frequency of vibration of rectangular wood and plywood plates ; and discussion 78, 487
Heat, Transient flow of, through a two-layer wall 481
Helmholtz line-element in brightness-colour space, A modified 41
Henderson, S. T., *see* Strange, J. W.
Herman, Renée, *see* Gaydon, A. G.
Herschel's condition 100
High-pressure combustion, The thermodynamic properties of the products of 737
Hopkins, H. G., *see* Smith-Rose, R. L.
Hopkins, H. H. : A transformer of known astigmatism formulae 663
Hopkins, H. H. : Herschel's condition 100
Hopkins, H. H. : The optical sine-condition 92
Horton, G. K., *see* Griffith, Miriam V.
Hydrostatic tension, The behaviour of water under 420, 436
Hysteresis and eddy losses in single crystals of an alloy of iron and silicon 21
Hysteresis of the angle of contact of mercury 120

Impedance of sound-absorbing materials, Apparatus for accurate measurement of acoustic; and discussion 253, 775
Initial temperature of cooling total radiator, The determination of ; and discussion 350, 356
Integral breadths of Debye-Scherrer lines produced by divergent x-rays 401
Interferometry, Low-order multiple-beam 654
International lumen and the lightwatt, The new 207
van Itterbeck, A. and van Doninck, W. : Velocity of sound in mixtures of argon, helium and hydrogen at low temperatures 615

Jones, D. E. (Obituary notice) 780

Klanfer, Laura, *see* Motz, H.
Klystron, Demonstration of a water-jet analogue of the reflection 475
Kudar, J. : Optical problems of the rotating-prism cinematograph projector 598

Lens systems, A system of transfer coefficients for use in the design of : VI 296
Lenses, aircraft camera, The performance of 493
Light-diffusing properties of magnesium oxide 408
Lightwatt, The new international lumen and the 207

PAGE

Linfoot, E. H. : Decentred aspheric plates . . . 65
Linfoot, E. H. : The Zernike phase-contrast test . . . 759
Lipson, H., *see* Farquhar, Margaret C. M.
Liquid, Dependence of pressure-variation with depth in a liquid on the vertical acceleration of the : demonstration . . . 118
Lord, Mary P. : Measurement of spectral distributions when lines and continua are present together . . . 477
Lord, Mary P. and Rees, A. L. G. : Behaviour of zinc sulphide phosphors under conditions of periodic excitation . . . 280
Lord, Mary P. and Rees, A. L. G. : Rapid determination of decay characteristics of luminescent solids . . . 289
Low temperatures, Velocity of sound in gas mixtures at . . . 615
Lumen, The new international, and the lightwatt . . . 207
Luminescence, Cathodo- : I–III . . . 369, 383, 392
Luminescence of solids . . . 247
Luminescent solids, The rapid determination of decay characteristics of . . . 289

Macfarlane, G. G., *see* Davies, H.
Magnesium oxide, The light-diffusing properties of . . . 408
Magnetic electron microscope, The variation of resolution with voltage in the . . . 443
Magneto-resistance of high-coercivity alloys . . . 153
Magnetron, Cavity-resonator . . . 247
Mailvaganam, A. W. : Tidal effects on the production of mesons in the atmosphere . . . 468
Manley J. J. (Obituary notice) . . . 332
Mechanical properties of metallic solid solutions . . . 669
Megaw, Helen D. : Crystal structure of double oxides of the perovskite type . . . 133
Mendelssohn's *z*-particles and the Bose-Einstein statistics . . . 302
Mercury, Hysteresis of the angle of contact of . . . 120
Mesons in the atmosphere, Tidal effects on the production of . . . 468
Metallic solid solutions, The mechanical properties of . . . 669
Motz, H. and Klanfer, Laura : Complete computation of electron optical systems . . . 30
Multiple-beam interferometry, Low-order . . . 654

Nabarro, F. R. N. : The mechanical properties of metallic solid solutions . . . 669
Networks dual to non-planar networks, Methods for the construction of . . . 677
Nitrogen, Band systems in the spectrum of . . . 292

Obituary notices . . . 332, 777
Oliver, R., *see* Ford, L. H.
Optical constants, The calculation of . . . 704
Optical glass absorption, Note on measurements of . . . 472
Optical phase-contrast test, Zernike . . . 759
Optical problems of the rotating-prism cinematograph projector . . . 598
Optical sine-condition . . . 92
Optical systems, Algebraic analysis of the higher-order aberrations of . . . 545
Optical systems, A system of transfer coefficients for use in the design of : VI . . . 296

Perovskite type, Crystal structure of double oxides of the . . . 133
Phosphors . . . 247
Phosphors, Behaviour of zinc sulphide, under conditions of periodic excitation . . . 280
Photochemical decomposition of CO; and discussion . . . 701
Photographic materials, Effect of the angle of incidence of the exposing light rays upon the resolving powers of . . . 769
Planckian (total) radiator, Cooling of a ; and discussion . . . 341, 350, 356
Porous material, The propagation of sound between walls of; and discussion . . . 358, 775
Potential curves and electronic states of diatomic molecules . . . 695
Pressure-variation with depth in a liquid, and vertical acceleration of the liquid : demonstration . . . 118

PAGE
Price, J. D. : Note on the calculation of optical constants 704
Projector, Optical problems of the rotating-prism cinematograph 598

Radar echoes from the sea surface at centimetre wave-lengths 717
Radiation of 9-cm. wave-length, Experimental investiagtion of the reflection and absorption of 265
Radiator, Cooling of a Planckian (total) ; and discussion 341, 350, 356
Radio sounding balloons and ultra-short-wave direction finding 184
Randall, J. T. : The cavity magnetron (and a note on the luminescence of solids) . 247
Receiving aerial, Fluctuation noise in a 313
Rectangular wood and plywood plates, Fundamental frequency of vibration of ; and discussion 78, 487
Rees, A. L. G., *see* Lord, Mary P.
Reflection and absorption of radiation of 9-cm. wave-length, Experimental investigation of the 265
Resolution, Variation with voltage in the magnetic electron microscope 443
Resolving powers of photographic materials, The effect of the angle of incidence of the exposing light rays upon the 769
Reviews of books 127, 210, 334, 488, 624, 781
Rupture, Extension of Griffith's theory of, to three dimensions 729
Rutherford : Life and work after the year 1919 (Second Rutherford Memorial Lecture) 625

Sack, R. A. : Extension of Griffiths' theory of rupture to three dimensions . . 729
Schmid, R. F. (Obituary notice) 333
Schmid, R. F. and Gerö, L. : Photochemical decomposition of CO; and discussion 701
Scott, R. A. : Absorption of sound in a homogeneous porous medium; and discussion 165, 775
Scott, R. A. : Apparatus for accurate measurement of the acoustic impedance of sound-absorbing materials; and discuwsion 253, 775
Scott, R. A. : Propagation of sound between walls of porous material; and discussion 358, 775
Scott, W. J. : Demonstration of a water-jet analogue of the reflection klystron . 475
Selwyn, E. W. H. and Tearle, J. L. : The performance of aircraft camera lenses . 493
Sine-condition, Optical 92
Single-crystal x-ray photographs, Accurate determination of cell dimensions from . 200
Single crystals of an alloy of iron and silicon, Hysteresis and eddy losses in . . 21
SiS, The rotational analysis of some near ultra-violet bands of 606
SiS, SiSe and SiTe, Ultra-violet band-systems of 538
Smith, T. : Note on measurements of (optical) glass absorption 472
Smith-Rose, R. L. and Hopkins, H. G. : Application of ultra-short-wave direction finding to radio sounding balloons 418
SnSe and SnTe, Ultra-violet absorption band-systems of 707
Solid solutions, The mechanical properties of metallic 669
Sound-absorbing materials, Apparatus for accurate measurement of the acoustic impedance of; and discussion 253, 775
Sound, Absorption of, in a homogeneous porous medium; and discussion . 165, 775
Sound, Propagation of, between walls of porous material; and discussion . 358, 775
Sound, Velocity of, in mixtures of argon, helium and hydrogen at low temperatures . 615
Spectra, Ultra-violet, of SiS, SiSe and SiTe. 538
Spectra, Ultra-violet absorption, of SnSe and SnTe 707
Spectra, vibrational, New regularities in 456
Spectral distribution of energy and the initial temperature of a cooling total radiator ; and discussion 350, 356
Spectral distribution when lines and continua are present together, Note on the measurement of 477
Spectrum of N_2, Band systems in the 292

PAGE

Spectrum of SiS, Rotational analysis of some near ultra-violet bands in the 606
Stiles, W. S.: A modified Helmholtz line-element in brightness-colour space 41
Stokes, Alexander R.: Construction and use of a "fly's-eye" for assisting x-ray structure analysis 306
Strange, J. W. and Henderson, S. T.: Cathodo-luminescence: I–III 369, 383, 392
Surface tension, Errors in viscometry due to 575
Surveys, Geomagnetic secular variations and 213

Tearle, J. L., *see* Selwyn, E. W. H.
Temperley, H. N. V. and Chambers, Ll. G.: The behaviour of water under hydrostatic tension: I and II 420, 436
Thermal conductivity and diffusion "constants" in non-homogeneous media, Measurement of 321
Thermal conductivity of textiles; and discussion 105, 305
Thermodynamic properties of the products of high-pressure combustion 737
Thermodynamic relations of rigid hygroscopic gels 585
Tidal effects on the production of mesons in the atmosphere 468
Tolansky, S.: Low-order multiple-beam interferometry 654
Total radiator, Cooling of a; and discussion 341, 350, 356
Total radiators, Colours of, expressed on the C.I.E. trichromatic system 1
Transfer coefficients for use in the design of lens systems: VI 296
Transformations of known astigmatism formulae 663
Transient flow of heat through a two-layer wall 481

Ultra-short-wave direction finding and radio sounding balloons 184

Vago, E. E. and Barrow R. F.: Ultra-violet absorption band-systems of SiS, SiSe and SiTe 538
Vago, E. E. and Barrow, R. F.: Ultra-violet absorption band-systems of SnSe and SnTe 707
Valatin, J. G.: Electronic states and potential curves of diatomic molecules 695
Vibration of rectangular wood and plywood plates, Fundamental frequency of; and discussion 78, 487
Vibrational spectra, New regularities in 456
Viscometry, Errors in, due to surface tension 575

Warburton, F. L.: Some thermodynamic relations of rigid hygroscopic gels 585
Water, The behaviour of, under hydrostatic tension 420, 436
Williams, E. J. (Obituary notice) 778
Wilson, A. J. C.: Hysteresis and eddy losses in single crystals of an alloy of iron and silicon 21
Wilson, A. J. C.: Integral breadths of Debye-Scherrer lines produced by divergent x-rays 401

X-ray photographs, Accurate determination of cell dimensions from single-crystal 200
X-ray structure analysis, Construction and use of a "fly's-eye" for assisting 306
X-rays, Integral breadths of Debye-Scherrer lines produced by divergent 401

Yarnold, The Rev. G. D.: The hysteresis of the angle of contact of mercury 120

z-particles, Mendelssohn's, and the Bose-Einstein statistics 302
Zernike phase-contrast test 759
Zinc sulphide phosphors, The behaviour of, under conditions of periodic excitation 280

INDEX TO REVIEWS OF BOOKS

PAGE
Baitsell, G. A. (Editor) : *Science in Progress* 129
Bergmann, S., *see* von Mises, R.
Boutry, G. A. : *Optique Instrumentale* 785
Brillouin, L. N. : *The Mathematics of Ultra-high Frequencies in Radio* 128
British Association Mathematical Tables, Vol. 1 : *Circular and Hyperbolic Functions, Exponential and Sine and Cosine Integrals, Factorial Function and Allied Functions, Hermitian Probability Functions* 339
British Association Mathematical Tables, Part-volume A : *Legendre Polynomials* . 783
British Association Mathematical Tables, Part-volume B : *The Airy Integral* . . 784
Bunn, C. W. : *Chemical Crystallography, an Introduction to Optical and X-ray Methods* 131

Committee for the Calculation of Mathematical Tables : *Table of Arc sin x*; *Tables of Associated Legendre Functions* 337

Daunt, J. G. : *Electrons in Action* 784

Frank, P. : *Thermodynamics* 338
Fletcher, A., Miller, J. C. P. and Rosenhead, L. : *An Index of Mathematical Tables* 491
Friedrichs, K. O., *see* von Mises, R.

Guinier, A. : *Radiocristallographie* 132

Heitler, W. : *Elementary Wave Mechanics* 127
Herzberg, G. : *Infra-red and Raman Spectra of Polyatomic Molecules* 489
His Majesty's Stationery Office ; *Polarographic and Spectrographic Analysis of High Purity Zinc and Zinc Alloys for Die Casting* 131

Le Grand, Y. : *Optique Physiologique* : *La Dioptrique de l'Œil et sa Correction* . . 624
Luneberg, R. K. : *Mathematical Theory of Optics* 335

Miller, J. C. P. : Mathematical Tables, Part-volume B : *The Airy Integral* . . 784
Miller, J. C. P., *see* Fletcher, A.
von Mises, R., Friedrichs, K. O. and Bergmann, S. : *Fluid Dynamics* 211

Newmark, M. : *Dictionary of Science and Technology in English-French German-Spanish* 781

Owen, D. : *Alternating Current Measurements at Audio and Radio Frequencies* . . 784

van der Pol, B. : *Music and Elementary Theory of Numbers* (The Music Review, Vol. VII, No.1 340
Prager, W. : *Mathematical Theory of Plasticity* 334

Robertson, J. K. : *Atomic Artillery and the Atomic Bomb* 338
Rosenhead, L., *see* Fletcher A.
Rothe, E. : *Questions Actuelles de Géophysique Théorétique et Appliquée* . . 128

Schrödinger, E. : *Statistical Thermodynamics* 340
Scott Blair, G. W. : *A Survey of General and Applied Rheology* 624
Sokolnikoff, I. S. : *Mathematical Theory of Elasticity* 488
Southwell, R. V. : *Relaxation Methods in Theoretical Physics* 782
Spreadbury, F. G. : *Electric Discharge Lighting* 782

Wilson, C. W. : *Radium Therapy. Its Physical Aspects* 210

8522